Ministry of Higher Education and Scientific Research

HassibaBenbouali University of Chlef Faculty of Technology

Process Engineering Department

Thesis

Submitted for the fulfillment of the degree of

ACADEMIC MASTER

Field: Process Engineering

Specialization: Pharmaceutical Engineering

Presented by:

GHERIOUNE Nesrine

BENMEHACHE Cherrouk

Thème :

Anti-malarial evaluation activity of some bioactive substances: an advanced computational approach leading to novel drug pre-formulation

Defended on 25/06/2025 before a jury composed of:

BENGHALIA Meryem Amal	MCA	University	President
OTMANINE Khaled	MCB	University	Supervisor
HENTABLI Mohamed	MCB	University	Examiner

We are thankful

First of all, we praise Allah Almighty for granting us the courage, will, and patience to complete this project, and for enabling us to carry it out with joy and determination.

We would like to express our sincere thanks and deep gratitude to our supervisor, Mr. Otmanine Khaled, for his valuable guidance, wise advice, and continuous support throughout the entire duration of this work. His trust in us and his professional insight were a great motivation to complete this work with care and dedication.

I would like to express my sincere thanks and deep appreciation to the esteemed members of the examination committee who honored me with their presence and evaluation of this work:

- **BENGHALIA Meryem Amal**, President of the committee, for her overall supervision and excellent leadership of the session.
- **OTMANINE Khaled**, Supervisor of this work, for his valuable scientific guidance and academic support throughout the preparation of this research.
- . **HENTABLI Mohamed**, Examiner, for his insightful remarks and scientific contributions that greatly enriched this work.

Please accept my utmost respect and heartfelt gratitude for your time, efforts, and support.

We also extend our heartfelt thanks to Dr. Hammoudi Mounir for his wisdom and support throughout this period.

Our appreciation goes as well to Mr. Bendriss Houari, whose efforts and encouragement greatly supported us during the realization of this research.

We are also deeply grateful to Professor Hanatabli Mohamed for his dedication, constant availability, and valuable assistance throughout the preparation of this work.

We would like to express our sincere gratitude to the SAIDAL pharmaceutical complex – Dar El Beïda for their cooperation and support, especially to Mrs. Mechied Wahiba for her availability and kindness, which made our training experience rich and fruitful.

Special thanks also go to Dalila and Wassila for their help, guidance, and warm welcome during our stay at the complex

Nesrine and cherrouk

Dedication

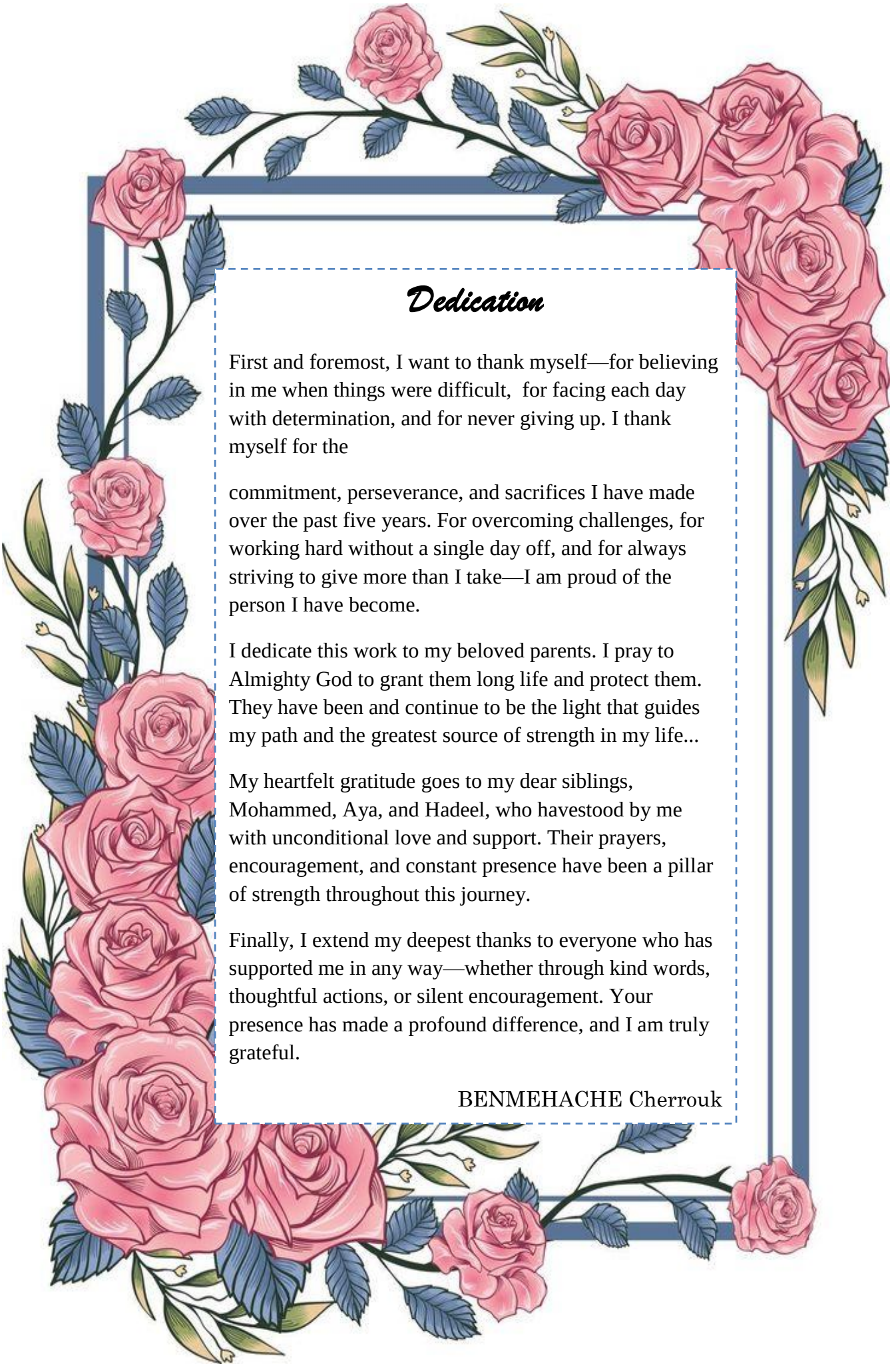
First and foremost, I want to thank **myself** — for believing in me when things were difficult, for showing up every single day, and for never giving up. I thank myself for the dedication, the perseverance, and the sacrifices I made over the past five years. For pushing through challenges, for working hard with no days off, and for always striving to give more than I receive — I am proud of the person I've become.

I dedicate this work to the memory of my beloved **late parents**. May Allah have mercy on their souls and grant them the highest ranks in Paradise. They were the light that guided my path and the greatest source of strength in my life. I pray that this work stands as an ongoing charity (sadaqah jariyah) for them, and that I continue to honor them through my efforts and achievements.

My heartfelt gratitude goes to my **dear siblings**, who have stood by me with unconditional love and support. Their prayers, encouragement, and constant presence have been a pillar of strength throughout this journey.

Finally, I extend my sincere thanks to everyone who supported me in any way—through kind words, thoughtful actions, or quiet encouragement. Your presence made a difference, and I am truly grateful

Gherioune Nesrine



Dedication

First and foremost, I want to thank myself—for believing in me when things were difficult, for facing each day with determination, and for never giving up. I thank myself for the

commitment, perseverance, and sacrifices I have made over the past five years. For overcoming challenges, for working hard without a single day off, and for always striving to give more than I take—I am proud of the person I have become.

I dedicate this work to my beloved parents. I pray to Almighty God to grant them long life and protect them. They have been and continue to be the light that guides my path and the greatest source of strength in my life...

My heartfelt gratitude goes to my dear siblings, Mohammed, Aya, and Hadeel, who havestood by me with unconditional love and support. Their prayers, encouragement, and constant presence have been a pillar of strength throughout this journey.

Finally, I extend my deepest thanks to everyone who has supported me in any way—whether through kind words, thoughtful actions, or silent encouragement. Your presence has made a profound difference, and I am truly grateful.

BENMEHACHE Cherrouk

SUMMARY :

Table List

Figure list

الملخص

تهدف هذه الدراسة إلى تقييم فعالية أربعة مركبات نباتية (أرتيميسينين من *الشيح الحولي*، تشامازولين من *الشيح الإفريقي*، والثيمول والكارفاكرول من *الزعتر الشائع*) كعوامل محتملة مضادة للملاريا. اعتمدت الدراسة على تقنيات المحاكاة الحاسوبية، مستهدفة أربعة بروتينات أساسية للطفيليات: PF-NDH2، بلازمبسين، PF-بلازمبسين_2، وبوليميراز الحمض النووي في الأبيكلوربلاست. أظهرت نتائج الالتحام الجزيئي (Molecular Docking) تفاعلات قوية ومواقع ارتباط محددة بدقة، لا سيما بين التشامازولين وPF-بلازمبسين_2 (ΔG = -7.7 kcal/mol) والأرتيميسينين و-PF NDH2 (ΔG = -7.4 kcal/mol). بالإضافة إلى ذلك، تم تطوير نموذج علاقة الكمية بين البنية والنشاط (QSAR) باستخدام الانحدار الخطي المتعدد (MLR) وانحدار متجهات الدعم (SVR)، استنادًا إلى مجموعة بيانات تضم 71 مشتقًا من المركبات المعروفة المضادة للملاريا. أظهر النموذج دقة تنبؤية عالية (2R>0.93 و RMSE<0.82).

الكلمات المفتاحية الملاريا . نبات الشيح الحولي . نبات الشيح الافريقي نبات الزعتر. التلاحم الجزيئي .

Abstract

This study aims to evaluate the effectiveness of four plant compounds as potential antimalarials: Artemisinin from *Artemisia annua* extract, Chamazulene from *Artemisia afra* extract, and both Thymol and Carvacrol from *Thymus vulgaris* (Thyme) extract. The study relied on computational simulation techniques, targeting four key parasite proteins: PF-NDH2, Plasmepsin, and PF-Plasmepsin_2, as well as Apicoplast DNA polymerase. Molecular docking results showed strong interactions and precisely defined binding sites, particularly between Chamazulene and PF-Plasmepsin_2 (ΔG = -7.7 kcal/mol) and Artemisinin with PF-NDH2 (ΔG = -7.4kcal/mol). In addition, a Quantitative Structure-Activity Relationship (QSAR) model was developed using Multiple Linear Regression (MLR) and Support Vector Regression (SVR), based on a dataset comprising 71 derivatives of known antimalarial compounds. The model demonstrated high predictive accuracy (R^2 > 0.93, and RMSE < 0.82).

Keywords: Malaria, *Artemisia annua*, *Artemisia afra*, *Thymus vulgaris*, Molecular Docking,QSAR.

List of Abbreviations:

2D: 2 dimensional

3D: 3 dimensional

ADME: Absorption, distribution, metabolism, and excretion

ApPol: Apicoplast DNA Polymerase

BBB: Blood-brainbarrier

DFT: DensityFunctionalTheory

DPPH: 1,1-diphenyl-2-picrylhydrazyl

HIA: Human intestinal absorption

HOMO: HighestOccupiedMolecular Orbital

IC50 : Concentration of inhibition 50

logP: The partition coefficient

LUMO: LowestUnoccupiedMolecular Orbital

MLR: Multiple LinearRegression

PF-NDH2: *Plasmodium falciparum*ubiquinoneoxidoreductase NDH2

PF-plasmepsin-2: *Plasmodium falciparum* of plasmepsin-2

Pf: *Plasmodium falciparum*

PLIP: Protein–ligand interaction profiler

QSAR: Quantitative Structure-Activity Relationship

RMSE: RootMean Square Error

SVM: Support Vector Machines

SVR: Support Vector Machines for Regression

UV: Ultraviolet radiation

WHO: World Health Organization

GeneralIntroduction

Introduction:

Malaria remains one of the oldest and most prevalent parasitic diseases, continuing to pose a significant global health threat. It is endemic in various regions, including South America, Africa, and tropical areas of Asia, with documented outbreaks also reported in North America and parts of Europe. Annually, approximately half a billion malaria cases are recorded, with the majority occurring in tropical regions of Asia and Africa **[1], [2]**.

This mosquito-borne disease is caused by protozoan parasites of the *Plasmodium* genus, primarily *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium* knowlesi , and *Plasmodium ovale*. Among these, *P. falciparum* is the most virulent species and a leading cause of severe malaria-related morbidity and mortality. However, *P. vivax* is also highly prevalent, particularly in Latin America, and is capable of causing life-threatening complications. *P. falciparum* primarily invades erythrocytes, utilizing host hemoglobin as a nutrient source **[3], [4], [5], [6]**.

This study embarks on an exciting journey to discover a ground breaking therapeutic approach. We will leverage advanced molecular modeling and screening techniques to pinpoint a novel process and a highly selective dual inhibitor. This innovative compound will simultaneously target key parasitic and host enzymes: **PF-NDH2**, **plasmepsins** (specifically **PF-plasmepsin_2**), and **Apicoplast DNA Polymerase**. Our ultimate goal is to develop a potent agent capable of combating both malaria and metabolic syndrome diseases. This research holds the promise of revolutionizing treatment paradigms by addressing critical pathways implicated in both conditions with a single, targeted intervention.

This research is divided into three main axes: molecular docking to study compound interactions with target proteins, QSAR analysis to model the structure-activity relationship of drug candidates, and experimental evaluation of compound efficacy. Through comprehensive investigation, three medicinal plants were identified as having the highest inhibitory effects against the parasite's key receptors: *Artemisia annua, Artemisia afra, and Thymus vulgaris.* This study combines advanced computational methodologies with practical applications to develop innovative anti-malarial treatments.

This comprehensive four-chapter study delves into the significance of **molecular modeling** and **molecular docking** as crucial computational tools in drug development. These techniques are employed to evaluate the interactions of small molecules with proteins and are further

augmented by **Quantitative Structure-Activity Relationship (QSAR) studies**, which link chemical structure to biological activity for predicting new compound properties. The research demonstrates the application of these advanced methodologies in extracting therapeutic compounds, such as **Artemisinin**, from the **Artemisia annua** plant, with the aim of developing innovative treatments for diseases like malaria and metabolic syndrome :

Chaptre I : Malaria is a severe infectious disease caused by *Plasmodium* parasites, particularly **Plasmodium falciparum**, and transmitted by Anopheles mosquitoes. It poses a global health challenge, especially for children and pregnant women. The disease's pathogenesis involves hepatic and erythrocytic stages; the latter leads to acute symptoms and severe complications like cerebral malaria and severe anemia due to red blood cell destruction and vascular occlusion. To combat increasing drug resistance, research focuses on identifying **new therapeutic targets** in the parasite's metabolic pathways and enzymes. Additionally, the exploration of **natural antimalarial compounds** such as **Artemisinin** from *Artemisia annua* , as they offer alternative mechanisms to fight the parasite.

Chaptre II : Molecular docking is a fundamental technique in drug development, crucial for new drug discovery due to its ability to significantly enhance the overall process efficiency [7]. **Scoring functions** are employed to assess how well small molecules bind to proteins, aiming to optimize binding affinity and therapeutic potential [8]. The advancement of **Molecular Modeling (MM)** marks a major breakthrough in computational chemistry, serving as a vital tool for identifying, designing, and optimizing new drug candidates [9, 10].

Chaptre III :Quantitative Structure-Activity Relationship (QSAR) studies employ chemometric methods to analyze how specific biological activity or physicochemical properties vary based on **molecular descriptors** that characterize a molecule's chemical structure. This approach allows for replacing expensive and time-consuming biological assays or experiments—especially those involving hazardous, toxic, or unstable compounds—with computationally derived descriptors [11, 12]. These descriptors can then be used to predict the relevant biological or physicochemical responses of novel compounds, leading to a more efficient and safer drug discovery and development process.

Chaptre IV : This work focuses on **extracting active components from *Artemisia annua***, a plant renowned for its wide-ranging therapeutic properties and long history in traditional medicine, particularly in the fight against malaria. The aim of this extraction process is to **isolate and concentrate pharmacologically important chemical compounds, such as**

Artemisinin. This paves the way for their evaluation and application as potential therapeutic agents, especially in the development of new drugs targeting diseases like malaria and metabolic syndrome.

Objectif:

In this study, we aim to find a solution to inhibit the growth of Plasmodium falciparum for the treatment of malaria using medicinal plants *(Artemisia annua, Artemisia afra, and thymusvulgaris*). Advanced drug discovery techniques, such as medicinal chemistry ,pharmacokinetics and Insilico studies , will be employed. The use of computational approaches in drug research enhances efficiency, effectiveness, and cost reduction, particularly through Quantitative Structure-Activity Relationship (QSAR) modeling. In this context, computational methodologies play a crucial role in identifying potential antimalarial drugs.

Problematic:

Is inhibiting proteins (apycoplast DNA plolymerase, plasmepsine, pf-plasmepsine-2, pf-NDH2) sufficient to treat malaria? Will artificial intelligence technology succeed in discovering a drug against malaria ?how can Quantitative Structure-Activity Relationship(QSAR) modeling guide the Design of new, improved treatments for

malaria?

Research Hypotheses :

1. Targeted Inhibition for Malaria Treatment:

- It is hypothesized that selectively inhibiting key Plasmodium falciparum proteins (PF-NDH2, plasmepsins , PF-plasmepsin-2, and Apicoplast DNA Polymerase) using a dual-action inhibitor will effectively halt parasite growth, leading to successful malaria treatment.

2. Efficacy of Medicinal Plant Extracts:

- Extracts from medicinal plants such as *Artemisia annua , Artemisia afra , and Thymus vulgaris* are expected to exhibit significant inhibitory effects against the parasite's key receptors.

3. Computational Drug Discovery Tools:

- Molecular modeling and screening tools (molecular docking and QSAR analysis) are anticipated to play a crucial role in identifying novel, selective anti-malarial inhibitors and efficiently predicting compound responses.

4. AI in Drug Discovery:

- Artificial intelligence (AI) and pharmacoinformatics are hypothesized to serve as innovative and successful approaches for discovering and developing new anti-malarial drugs, improving efficiency while reducing costs.

Chapter I : Malaria: Epidemiology, Pathogenesis, and Potential Therapeutic Targets in Plasmodium falciparum with Focus on Natural Antimalarial Compounds

I)-Malaria:

1. Background:

Malaria remains a significant global health concern, particularly in tropical and subtropical regions where transmission is endemic. Currently, 91 countries and territories report ongoing malaria transmission, with over 125 million international travelers visiting these regions annually. A considerable number of these travelers acquire malaria while in endemic areas, with more than 10,000 cases documented among returning travelers each year; however, due to underreporting, the actual incidence is likely much higher**.**

International travelers originating from malaria-free regions and visiting endemic areas are at high risk of infection due to a lack of acquired immunity. Similarly, migrants from malaria-endemic regions residing in non-endemic countries are at increased risk when returning to visit friends and relatives, as any previously acquired immunity may have waned or been lost over time.

Travelers who contract malaria during their journey often face barriers to timely diagnosis and treatment, including limited access to reliable healthcare services. Upon returning to malaria-free regions, they may encounter delays in diagnosis due to lack of clinical awareness among healthcare providers. Additionally, the unavailability or lack of registration of effective antimalarial drugs in non-endemic countries can contribute to disease progression, increasing the risk of severe complications and mortality.

Any traveler presenting with febrile illness within three months of returning from a malaria-endemic area should be evaluated immediately to rule out malaria, as delayed diagnosis can result in life-threatening complications. In settings where rapid diagnostic testing is unavailable, standby emergency treatment (SBET) may be warranted as a precautionary measure **[13].**

2.Geographic Risk and Preventive Measures:

Travelers to forested regions of Southeast Asia, including Brunei Darussalam, Cambodia, China, Indonesia, Lao PDR, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam, should be aware of the risk of *P. knowlesi* infection. Protective measures, including vector control (e.g., insect repellent, bed nets, and protective clothing) and chemoprophylaxis where indicated, are essential to prevent infection**[13].**

3. Geographical Distribution of Malaria:

The global distribution of malaria is continuously monitored and reported in the World Malaria Report by the World Health Organization (WHO). The risk of malaria transmission for travelers varies significantly between countries and even within different regions of the same country, necessitating careful assessment when considering preventive measures.

In most malaria-endemic regions, large urban centers are typically free from malaria transmission, though peri-urban areas may still pose a risk. However, malaria is actively transmitted in urban environments across sub-Saharan Africa and, to a lesser extent, in India**[13] (see figure 1).**

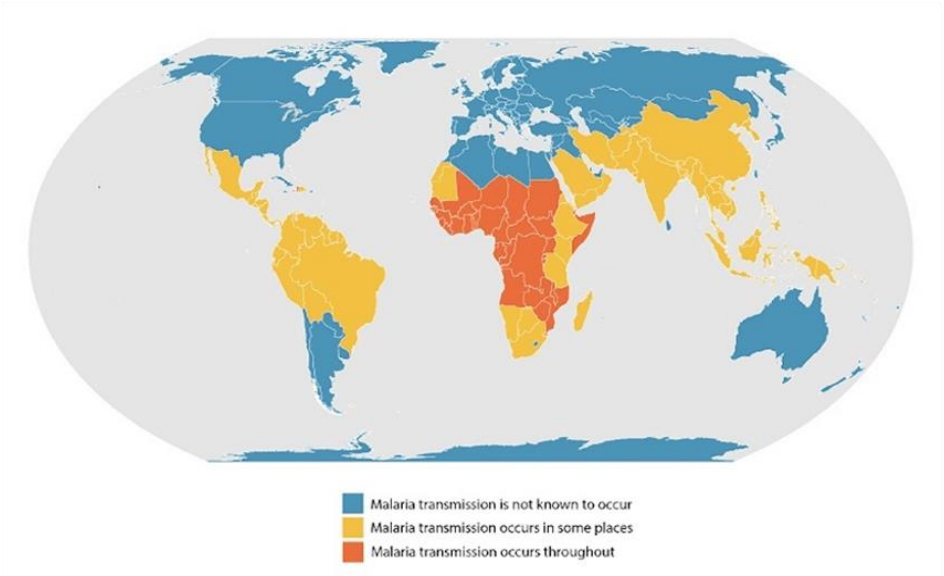


Figure 1: the spread of malaria in the world**[15]**

4.Geographical and environmentalfactors influence malaria transmission:

- Altitude: The risk is generally lower at altitudes above 1,500 meters, but in favorable climatic conditions, transmission can occur at elevations approaching 3,000 meters.
- Seasonality: The highest transmission risk typically occurs at the end of or shortly after the rainy season.

Certain tourist destinations in Southeast Asia, the Caribbean, and Latin America are free of malaria risk. For precise, country-specific malaria risk assessments, refer to the Country List provided in WHO reports**[13].**

New data from the WHO reveal that an estimated 2.2 billion cases of malaria and 12.7 million deaths have been averted since 2000, but the disease remains a serious global health threat, particularly in the WHO African Region. According to WHO's latest *World malaria report*, there were an estimated 263 million cases and 597 000 malaria deaths worldwide in 2023. This represents about 11 million more cases in 2023 compared to 2022, and nearly the same number of deaths. Approximately 95% of the deaths occurred in the WHO African Region, where many at risk still lack access to the services they need to prevent, detect and treat the disease**[14].**

5. Etiology of Malaria:

Malaria is a vector-borne disease caused by protozoan parasites of the genus Plasmodium. In humans, infection is attributed to five primary species: Plasmodium falciparum, Plasmodium malariae, Plasmodium ovale, Plasmodium vivax, and Plasmodium knowlesi.

Among these, P. falciparum and P. vivax are the most prevalent globally, with P. falciparum being the most virulent species, responsible for the highest morbidity and mortality rates. Severe P. falciparum malaria is a major public health burden, particularly in sub-Saharan Africa, where it accounts for the majority of malaria-related fatalities.

Although P. knowlesi primarily infects non-human primates, zoonotic transmission to humans has been documented. However, there is currently no evidence of human-mosquito-human transmission of P. knowlesi or other zoonotic malaria species. Further research is required to assess the potential risk of sustained transmission in human populations**[13].**

6. Transmission:

Malaria is transmitted through the bite of an infected female Anopheles mosquito, which serves as the primary vector for Plasmodium spp. Transmission occurs predominantly between dusk and dawn, as Anopheles mosquitoes exhibit nocturnal feeding behavior. During a blood meal, the mosquito injects sporozoites into the human host, initiating the infection cycle. The efficiency of transmission is influenced by various factors, including mosquito species, environmental conditions, and human-vector interactions**[13].**

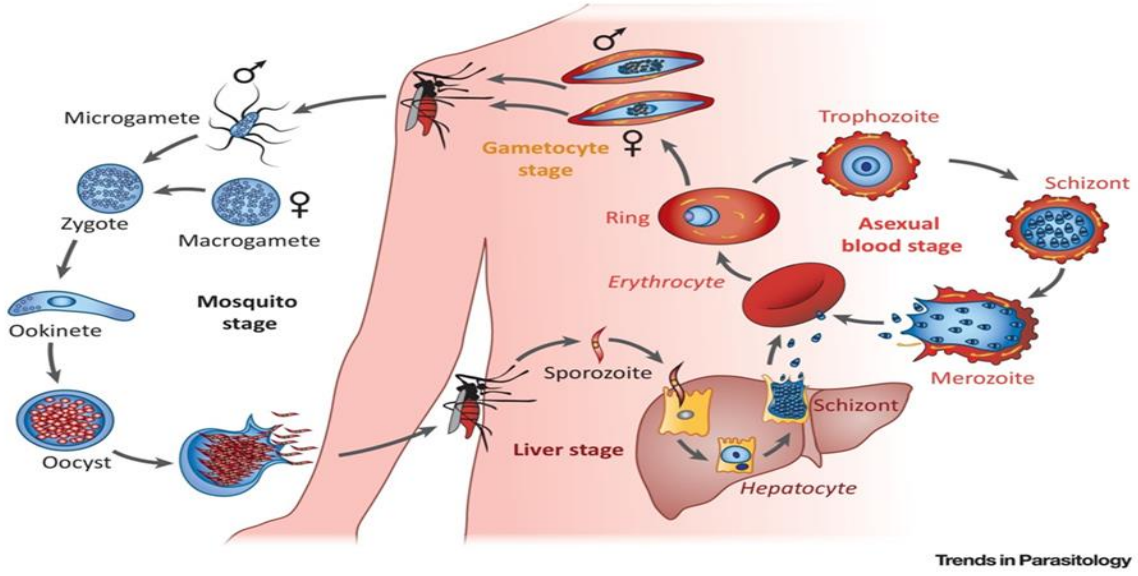


Figure 2: The figure illustrates the life cycle of the malaria parasite (Plasmodium), which causes malaria and is transmitted through the bite of an infected female Anopheles mosquito**[16].**

7.The life cycle is divided into three main stages:

7.1. Liver Stage (Hepatic Stage)

- When an infected mosquito bites a human, it injects the parasite in the form of sporozoites into the bloodstream.
- The sporozoites travel to the liver, where they infect liver cells (hepatocytes) and develop into schizonts.
- The schizonts multiply and burst, releasing merozoites into the bloodstream.

7.2. Asexual Blood Stage

- The merozoites infect red blood cells (erythrocytes) and develop into ring-stage trophozoites.
- These trophozoites grow and mature into schizonts, which multiply and cause the red blood cells to rupture.
- This releases more merozoites, which continue infecting new red blood cells, causing cycles of fever and chills in malaria patients.

7.3. Sexual Stage & Transmission ThroughMosquitoes

- Some parasites differentiate into male and female gametocytes inside red blood cells.
- When a mosquito bites an infected person, it ingests these gametocytes.
- Inside the mosquito's gut, microgametes (male) and macrogametes (female) fuse to form a zygote, which then develops into an ookinete and later into an oocyst.
- The oocyst matures and bursts, releasing new sporozoites that migrate to the mosquito's salivary glands, making it ready to transmit malaria when it bites another person.

7.4.Clinical Characteristics of Malaria:

Malaria is an **acute febrile illness** with an **incubation period of at least seven days** following exposure. Consequently, any febrile illness occurring **one week or more** after potential exposure to malaria should prompt clinical suspicion.

8.Severe Malaria and Complications

The most severe form of malaria is caused by *Plasmodium falciparum*, which is associated with a wide spectrum of clinical manifestations, including:

- Non-specific symptoms: Fever, chills, headache, myalgia, weakness, vomiting, cough, diarrhea, and abdominal pain.
- Severe complications: In cases of disease progression, organ failure may develop, leading to acute renal failure, pulmonary edema, generalized convulsions, circulatory collapse, coma, and death.

Due to the non-specific nature of early symptoms, malaria can be misdiagnosed as other endemic febrile illnesses such as dengue fever, acute respiratory infections, and septicemia. Given the potential severity of *P. falciparum* malaria, any unexplained fever occurring between seven days and three months (or, rarely, later) after exposure should be urgently investigated. Prompt diagnosis and treatment are critical, as delays beyond 24 hours after symptom onset significantly increase the risk of fatal outcomes.

Certain groups are at **higher risk** of severe malaria, including:

- **Young children**
- **Pregnantwomen**
- **Elderlyindividuals**
- **Immunosuppressedindividuals**

Malaria in pregnant women—especially *P. falciparum* infection in non-immune individuals—substantially increases the risk of maternal mortality, miscarriage, stillbirth, and neonatal death.

Malaria Caused by Other *Plasmodium* Species:

- *P. vivax* and *P. ovale* can establish latent hepatic infections through hypnozoites, leading to relapses months or even years after initial exposure. *Primaquine* is currently the only drug capable of eliminating hypnozoites and preventing relapses.
- *P. malariae* can persist in the bloodstream for years, often asymptomatically, but rarely causes life-threatening disease.
- *P. knowlesi*, a zoonotic malaria parasite primarily affecting non-human primates, has been increasingly reported in forest-associated regions of Southeast Asia. Although human-mosquito-human transmission has not been documented, sporadic cases have been observed in travelers. Severe *P. knowlesi* malaria can lead to organ failure and fatal outcomes, but unlike *P. vivax* and *P. ovale*, it does not have persistent liver stages, and relapses do not occur**[13](see annex).**

We aim for selective inhibition targeting key enzymes and proteins in Plasmodium falciparum, the parasite responsible for malaria, by focusing on:

PF-NDH2 ;Plasmepsins ; PF-Plasmepsin_2 ; Apicoplast DNA Polymerase

This strategy aims to disrupt critical metabolic pathways in P. falciparum while minimizing effects on human cells, making it a promising approach for developing new antimalarial therapies

Presentation of receptors :

1.Apicoplaste Dna Polymerase :

The apicoplast harbors its own genome, necessitating a specialized set of DNA replication and repair enzymes. Among these, the apicoplast DNA polymerase (apPol) plays a crucial role in the replication of the apicoplast's circular DNA. Notably, apPol is the sole DNA polymerase localized to the apicoplast, making it a promising target for drug development. As a member of the A-family DNA polymerases, apPol shares functional characteristics with other well-studied polymerases involved in genomic replication and repair across various organisms. However, it likely exhibits unique structural and functional adaptations that reflect its specialized role and the evolutionary trajectory of the apicoplast**[17].**

Phylogenetic analyses suggest that apPol is related to archaealMethanomicrobia within the phylum Euryarchaeota, as well as to polymerases from diverse bacterial phyla. Furthermore, apPol retains the conserved motif B3 phenylalanine, a feature shared with DNA Polymerase I and Plant Organellar DNA polymerases. The apo structure of apPol has been resolved, and its biochemical properties have been characterized, indicating that apPol exhibits high fidelity but low processivity during DNA synthesis. However, the dynamic interactions between apPol, its DNA substrate, and incoming nucleotides during catalysis remain poorly understood due to the absence of DNA-bound structural data**[17]**

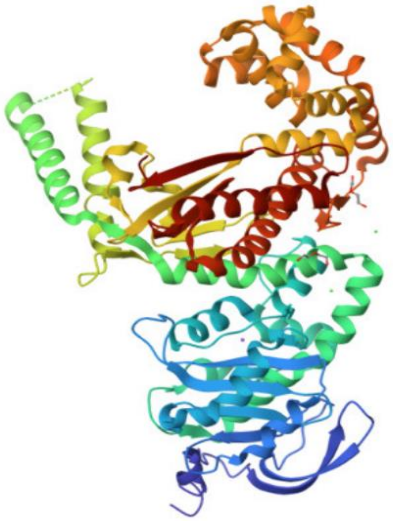


Figure 3: Apicoplaste Dna Polymerase (apPol)**[18]**

The apicoplast possesses an independent 35 kb genome and harbors approximately 600 proteins, the majority of which are encoded by nuclear genes of the parasite [19]. Among the various proteins involved in apicoplast DNA replication, the apicoplast DNA polymerase (apPOL) is essential for parasite survival and exhibits a clear prokaryotic origin. Its closest homolog outside the Apicomplexa is the replicative polymerase from the cyanobacterium*Cyanothece sp. PCC 8802*, sharing 35% protein sequence identity **[20], [21], [22].** In contrast, the most similar human polymerases, the lesion bypass polymerases theta and nu, exhibit only 23% and 22% sequence identity, respectively, highlighting the potential for selective targeting. Additionally, apPOL from *Plasmodium falciparum* and *Plasmodium vivax* share 84% sequence identity, suggesting that inhibitors designed against Pf-apPOL could also be effective against *P. vivax*, reinforcing its potential as a promising target for novel antimalarial drug development **[23] [24].**

2. PF-NDH2 :

Type II NADH:ubiquinoneoxidoreductase (NDH2) is a component of the mitochondrial electron transport chain that is absent in mammals but present in various organisms, including the malaria parasite. NDH2 has been proposed as a potential antimalarial drug target for several years. However, recent findings challenge its essentiality, as PfNDH2 was shown to be dispensable during the asexual blood stages of *Plasmodium falciparum*development**[25].** Conversely, in the *Plasmodium berghei* mouse model, NDH2 has been demonstrated to be

crucial for sporogony within the mosquito vector, suggesting its potential role in malaria transmission**[26].**

The mitochondrial type II NADH dehydrogenase (NDH2) of *Plasmodium falciparum*, PfNDH2 (PF3D7_0915000), has been recognized as a promising antimalarial drug target for over a decade. Unlike mammalian cells, which possess a conventional multi-subunit NADH dehydrogenase (Complex I) within the mitochondrial electron transport chain (mtETC), *Plasmodium* parasites rely on a single-subunit NDH2. This enzyme lacks proton-pumping activity and has no human homolog, making it an attractive candidate for selective therapeutic intervention**[27].**

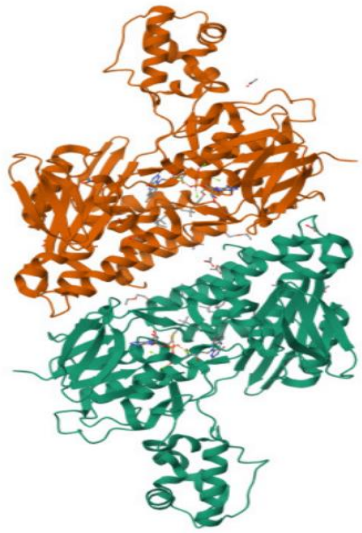


Figure4: plasmodium falciparum ubiquinone oxidoreductase NDH2 (PF-NDH2)**[28]**

3.Plasmepsins:

Plasmepsins are a diverse family of aspartic proteases in the malaria parasite *Plasmodium*, exhibiting a wide range of functions critical to parasite survival. These roles include hemoglobin degradation, processing of secretory organelle proteins essential for egress and invasion, and effector export. While certain plasmepsins, particularly those associated with the digestive vacuole, have been extensively studied, others, such as those involved in transmission stages, remain poorly characterized **[29].**

Plasmepsins are members of the ancient A1 aspartic protease family, also known as the pepsin-like family, which is widely distributed among eukaryotes. Among the ten identified plasmepsins, the digestive vacuolar plasmepsins (PM I–IV) are the most closely related.

These enzymes are encoded within a compact 16-kilobase region of chromosome 14 and share 50–70% amino acid sequence identity. Beyond *Plasmodium falciparum* and other primate-infecting *Plasmodium* species, the distribution and functional diversity of plasmepsins remain less well characterized**[29].**

Figure 5: plasmepsins**[30]**

4.Plasemoduim falciparum of plasmepsin II:

Plasmepsin II (Plm II) exhibits characteristic features of aspartic proteases, resembling enzymes such as cathepsin D, renin, and HIV protease**[31],[32]**. The mature Plm II consists of 329 amino acid residues and adopts a single-chain structure, folding into two topologically similar N-terminal and C-terminal domains. As observed in other aspartic proteases, the apo conformation of Plm II retains a catalytic water molecule within its active site, which facilitates peptide bond cleavage by nucleophilic attack on the carbonyl carbon of the catalytic aspartic dyad. Structural analysis reveals that this catalytic water molecule is positioned at distances of 2.84 Å and 2.75 Å from the oxygen atoms of the dyad's carboxyl groups**[32].**

The N-terminal domain of Plm II features a distinctive β-hairpin structure, referred to as the flap, which plays a critical role in inhibitor recognition and binding, akin to the mechanism observed in HIV protease. The structural dynamics of Plm II involve an initial flap opening to accommodate the inhibitor, followed by hydrogen bond formation or rigidification, ensuring stable inhibitor retention within the active site. A key residue, Leu292, located within the flexible proline-rich loop (Ile290–Pro297), collaborates with the N-terminal

flap to facilitate inhibitor entry through an open conformation, subsequently transitioning into a closed conformation to secure the inhibitor within the active site. These structural insights are instrumental in guiding the rational design of plasmepsin II inhibitors**[33],[31],[32][6].**

Figure 6: plasmodium falciparum of plasmepsin-2**[34]**

Over the past fifty years, natural compounds, particularly those derived from plants, have played a crucial role in pharmaceutical advancements, significantly contributing to drug discovery and development. Numerous pharmacological classes, originally based on bioactive plant-derived compounds, have proven effective in treating a wide range of infectious diseases, cancer, and metabolic disorders. Notably, quinine and its synthetic derivatives, which are fundamental to antimalarial therapy, are derived from plant sources**[8].**

5.Artemisia plant

The genus *Artemisia* L., one of the largest within the Asteraceae family, comprises over 500 species predominantly distributed across the northern temperate regions. Many species have been traditionally utilized in folk medicine for various therapeutic applications since ancient times**[35].** Among them we mention two type of artimisiaplants :*Artemisia annua and Artemisia afra*

5.1.*Artimisia annua* (الشيح الحولي):

Artemisia annua belongs to the class *Magnoliopsida*, which encompasses flowering plants The *Asteraceae* family is the second-largest family of flowering plants globally. *Artemisia* is a diverse and extensive genus, comprising between 200 and 400 species, including hardy herbs and shrubs. *A. annua* is an annual shrub that grows to a height of 50–150 cm. It thrives in temperate climates and is predominantly found in China and Vietnam, while also being cultivated in East Africa, the United States, Russia, and India**[36].**

Figure7: Plants of *artemisia annua***[37].**

The discovery and development of artemisinin (ART) from *Artemisia annua* L. (*Asteraceae*) have led to the emergence of a novel class of highly effective antimalarial agents, significantly impacting malaria treatment .Artemisinin-based combination therapies (ACTs) are currently regarded as the most effective treatment for uncomplicated *Plasmodium falciparum* malaria Moreover, ACTs are unlikely to be compromised by resistance in the near future, provided widespread access to this therapy is ensured **[36].**

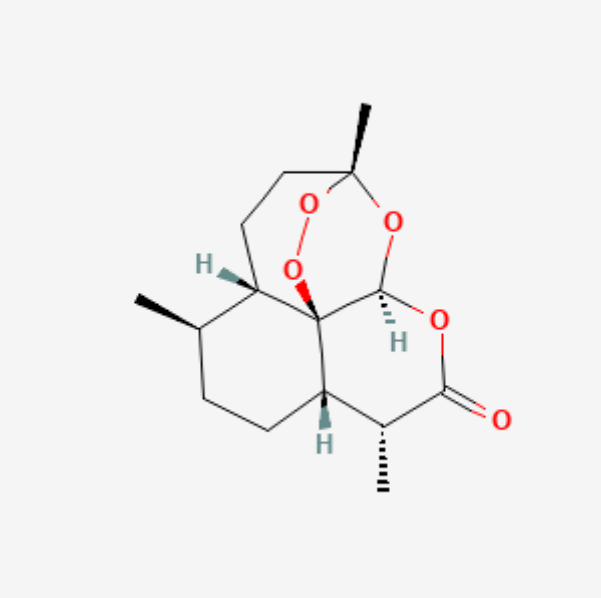


Figure8: Chemical Structure of Artemisinin**[38]**

Artemisia annua is recognized as a significant ethnomedicinal plant, with its most common mode of administration being a whole-plant decoction, traditionally used for the treatment of colds, coughs, and malaria. Additionally, powdered dried leaves have been employed in managing diarrhea. The entire plant exhibits various pharmacological properties, including antiseptic, antipyretic, anthelmintic, antispasmodic, and stimulant activities **[39].**Furthermore, crushed *A. annua* has been utilized in the preparation of liniments, while its tincture has been applied in the treatment of nervous disorders . In several African countries, tea infusions derived from *A. annua* are traditionally used for malaria treatment**[38].**

Artemisia annua: **A Traditional Chinese Medicinal Herb with Antipyretic and Antimalarial Properties**

Artemisia annua is a widely utilized herbal medicine in Traditional Chinese Medicine (TCM), characterized by its bitter taste and cold nature. It is associated with the hepatic, biliary, and renal meridians and is traditionally used for its effects in clearing summer heat, alleviating febrile conditions, and exhibiting antimalarial properties. Clinically, *A. annua* has been employed in the management of fever due to Yin deficiency, heat-related exogenous afflictions, malaria, and jaundice caused by damp-heat accumulation. The therapeutic applications and pharmacological properties of *A. annua* have been extensively documented in historical texts on Chinese materiamedica**[40].**

5.2.*Artemisiaafra* (الشيح الافريقي):

Artemisia afra,commonly known as African wormwood, is referred to by various names across different cultures **[41].** Native to Africa, *A. afra* is widely distributed across the continent and has been identified in various regions worldwide due to its significant medicinal potential. This woody perennial herb can grow up to at least 2 meters in height. It is characterized by finely divided, oval-shaped leaves, an aromatic fragrance, and alternately arranged silver-grey leaves on the adaxial side, while the abaxial side exhibits a lighter green coloration **[42][43].**

Figure9:plants of *artimisia afra**[*44**]***

Artemisia afra has a longstanding history of ethnomedicinal use in both the northern and southern regions of Ethiopia. In the Bale area, a juice prepared from chopped leaves mixed with water is traditionally administered orally for the treatment of roundworm infections and stomach pain. Additionally, the leaves are chewed, or their aroma is inhaled, to alleviate stomachaches and headaches**[45]**. In northern Ethiopia, charred leaf powder mixed with honey or edible oil is used as a remedy for eye diseases, including excessive tearing and cataracts, as well as stomach cramps.

Furthermore, tea infusions prepared from the leaves are commonly employed for the treatment of coughs, colds, flu, and bronchial and intestinal ailments. Decoctions of the whole plant or leaves in milk are utilized in managing haematuria, hemorrhoids, mumps, smallpox, malaria, neuralgia, colitis, and liver disorders. *A. afra* is also known for its tonic, stimulant,

aromatic, anthelmintic (vermifuge), and antipyretic (febrifuge) **[46],[47],[48],[49].**Among the antimalarial molecules of *A.afra* :Chamazulene

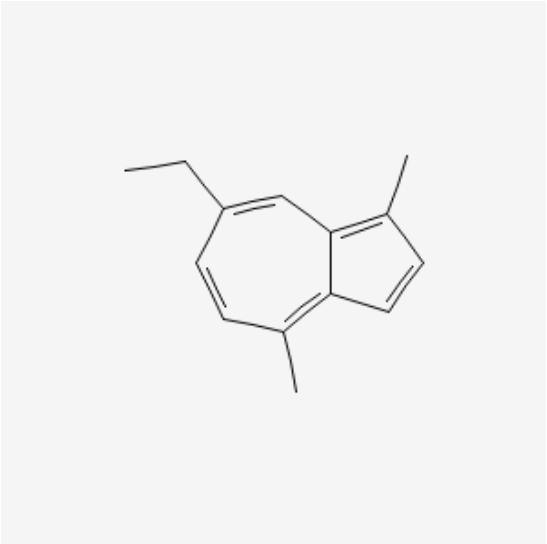


Figure 10: chemical structure of chamazulene**[50]**

Artemisia annua, recognized in the Chinese Pharmacopeia, is traditionally employed for the treatment of malaria, fever, and various other ailments. Similarly, Artemisia afra, as documented in the African Pharmacopeia, is utilized for the management of rhinitis, dyspepsia, cough, malaria, and renal disorders. Notably, herbal infusions derived from both *A. annua* and *A. afra* exhibit significant therapeutic potential in the regulation of diabetes mellitus. The bioactive compounds primarily associated with their antidiabetic properties include flavonoids and essential oils **[51].**

6.*Thymus vulgaris (الزعتر)* :

Thymus vulgaris, commonly known as thyme, is a flowering plant belonging to the Lamiaceae family. It is native to Southern Europe and has a widespread global distribution **[52].** The species is indigenous to the Mediterranean region and adjacent areas, including Northern Africa and parts of Asia. In Africa, thyme is cultivated in several countries, including Egypt, Morocco, Algeria, Tunisia, and Libya **[53][54].**

Figure 11: Plants of *thymus vulgaris*[**55]**

Thymus vulgaris has been investigated for its potential therapeutic applications beyond its common use as a culinary herb. Studies have highlighted its anti-inflammatory, antibacterial, antifungal, antiviral, antioxidant, and antispasmodic properties, demonstrating its potential in various medicinal applications**[56].**

Thymus vulgaris is a herbaceous plant widely distributed in southern Europe and commonly known as thyme**[54].** Herbs belonging to the Lamiaceae family are among the most common sources of thymol in their essential oils, with *Thymus vulgaris* being one of the most prominent plants containing this compound. Thymol (2-Isopropyl-5-methylphenol) is a naturally occurring monoterpene phenol, derived from cymene, and an isomer of carvacrol**[57].**

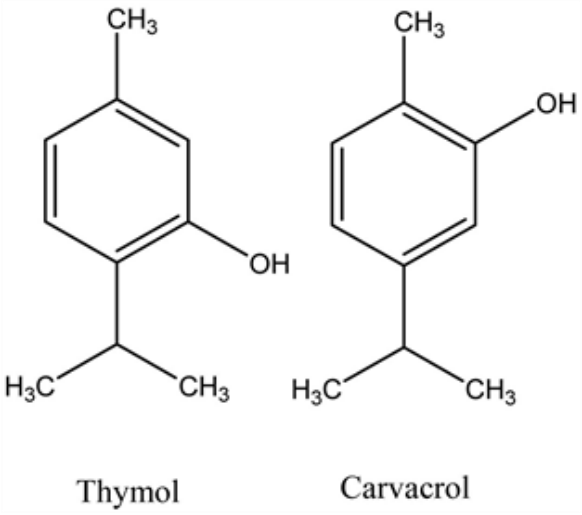


Figure 12: chemical structure of carvacrol and thymol**[57]**

Conclusion:

Malaria continues to pose a major global health threat, particularly in tropical regions where Plasmodium falciparum drives high mortality rates. The increasing resistance to existing antimalarial drugs highlights the need for innovative treatments targeting essential parasite pathways, such as apicoplast DNA polymerase, PF-NDH2, and plasmepsins, which are crucial for parasite survival and offer potential for selective drug development. Natural compounds, particularly artemisinin from*Artemisia annua* and chamazulene from *Artemisiaafra*, have proven highly effective and form the basis of current therapies. Additionally, plant-derived molecules like thymol and carvacrol from *Thymus vulgaris* show promising antimalarial properties. Moving forward, research should focus on optimizing drug design, developing combination therapies to combat resistance, and exploring traditional medicinal plants for new antimalarial leads. Combining modern molecular approaches with ethnobotanical knowledge could pave the way for more effective and sustainable malaria treatments, bringing the world closer to eradicating this deadly disease.

Chaptre II : Molecular Docking: Principles, Tools, and Applications in Drug Discovery

Introduction

has proven to be an extremely useful technique in various areas, including clinical training management, sequence analysis platforms, and molecular modelling. The development of new drugs relies heavily on in silico techniques, particularly molecular docking, due to the complex and multistep nature of the process. Molecular docking enhances the efficiency and effectiveness of the entire drug development pipeline**[7].**

1. Principle of Molecular Docking:

Molecular docking is a computational method employed to predict the interaction between a ligand and the active site of a target protein in three-dimensional (3D) space. This technique is fundamental for elucidating the mechanisms of molecular recognition and is pivotal in the drug discovery process.

The docking procedure encompasses several essential steps, including the prediction of binding affinity, accurate identification of binding sites, and determination of the ligand's binding mode. It also accounts for the diversity of ligands and the variability in their orientations to enhance predictive accuracy.

Overall, molecular docking provides a powerful framework for modeling ligand-protein interactions, estimating binding affinities, and clarifying the structural basis of molecular recognition. It is extensively utilized in virtual screening and structure-based drug design to facilitate the development of novel bioactive compounds **[59]**.

2. The Role of Optimization (Density Functional TheoryDFT):

Geometry optimization using Density Functional Theory (DFT) is a computational technique aimed at determining the most stable atomic configuration of a molecule or material

The primary objective is to identify the atomic arrangement corresponding to the system's minimum energy state, commonly referred to as the ground state **[58].**

In DFT-based geometry optimization, the atomic positions within the molecular or material structure are iteratively adjusted to minimize the total energy of the system. This procedure involves solving the Schrödinger equation within the framework of density functional theory to describe the electronic structure. At each iteration, atomic positions are slightly modified, and the system's total energy is recalculated to guide the structure towards the minimum.

Efficient optimization algorithms, such as the conjugate gradient method and the quasi-Newton method, are commonly employed to expedite convergence to the minimum energy configuration **[58].**

Upon completion of the optimization process, the obtained geometry represents a local minimum on the potential energy surface, corresponding to the most stable atomic arrangement under the given computational parameters. This optimized structure can then be utilized for subsequent studies, including the prediction of molecular properties, simulation of spectroscopic features, and investigation of chemical reactivity **[58][59]**.

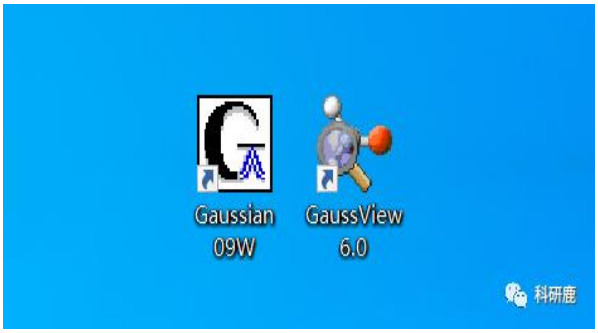


Figure 13 : Gaussian view optimization (Density Functional Theory DFT)

3.Bioinformatics Tools and Databases :

A)-Pubchem :

PubChem is a publicly accessible repository that provides comprehensive information on chemical substances and their biological activities. Launched in 2004 as part of the U.S. National Institutes of Health (NIH) Molecular Libraries Roadmap Initiative, PubChem has since evolved into a critical resource for researchers in diverse fields, including cheminformatics, chemical biology, medicinal chemistry, and drug discovery.

Each database uses a distinct primary identifier: Substance ID (SID) for the Substance database, Compound ID (CID) for the Compound database, and Assay ID (AID) for the BioAssaydatabase**[60]**

B)-Pdb:

The Protein Data Bank (PDB;) serves as the central international archive for three-dimensional structural data of biological macromolecules. This publication outlines the objectives of the PDB, its data deposition and retrieval systems, and methods for accessing

additional resources. Furthermore, it highlights the short-term development strategies aimed at enhancing this essential scientific resource**[61].**

C) AutoDockVina Tools:

AutoDock, along with its associated tools, plays a crucial role in structure-based drug design, virtual screening, and the analysis of molecular interactions. It offers researchers robust computational methods for studying ligand-receptor interactions and predicting potential binding modes of candidate drugs **[62].**

In this step, the ligand and target files are prepared by converting them from PDB (Protein Data Bank) format to PDBQT format, which is required for docking simulations using AutoDockVina**[58].**.

D)-AutoDockVina:

AutoDockVina, developed by Dr. Oleg Trott in 2010 at the Molecular Graphics Laboratory of The Scripps Research Institute, represents a major advancement within the AutoDock suite. This molecular docking software provides enhanced efficiency, ease of use, and greater accuracy in predicting ligand-receptor interactions. Its intuitive interface, combined with advanced optimization algorithms, establishes AutoDockVina as a powerful tool for structure-based drug design and virtual screening applications **[62][59].**

E)- Databases:

- **CB-Dock:**

CB-Dockis a versatile and efficient platform designed for protein-ligand blind docking, enabling accurate prediction of ligand binding poses and analysis of binding sites. It achieves a success rate exceeding 70% for top-ranked poses with minimal root-mean-square deviation (RMSD) values, providing an intuitive and effective environment for identifying protein binding sites without requiring prior structural knowledge.

Furthermore, CB-Dock2, an enhanced version of the original tool, demonstrates superior performance compared to previous blind docking methods, achieving a binding pose prediction success rate of approximately 85%. This advancement is attributed to the integration of CB-Dock with a template-based docking engine, thereby improving both accuracy and reliability **[59].**

- **SwissADME:**

generates a graphical output integrating the BOILED-Egg model, as originally proposed, to predict passive diffusion related to human intestinal absorption (HIA) and blood-brain barrier (BBB) permeability. This approach visualizes the molecular distribution within the WLOGP versus TPSA physicochemical space. In this analysis, SMILES representations of the compounds, retrieved from PubChem, are submitted to SwissADME to assess their hydrophilic or lipophilic characteristics according to the BOILED-Egg model **[63][59].**

- **SwissTarget Prediction:**

is an online tool developed to predict the probable biological targets of bioactive small molecules in humans and other vertebrate species. It supports the elucidation of molecular mechanisms underlying specific phenotypes or bioactivities, aids in the rationalization of potential side effects, and facilitates the prediction of off-target interactions for known compounds **[64][59].**

- **ProTox 3.0 :**

is a web-based virtual toxicity laboratory intended for academic and non-commercial use. It enables the prediction of various toxicological endpoints based on the chemical structure of compounds. The platform employs computational models trained on extensive experimental data (both in vitro and in vivo) to assess the toxicological potential of both existing and virtual compounds.

ProTox 3.0 predicts acute toxicity classes and a range of toxicological endpoints by analyzing chemical similarity to known toxic compounds and utilizing machine learning algorithms. As an openly accessible in silico tool, it offers a comprehensive resource for toxicity prediction, serving toxicologists, regulatory agencies, computational chemists, and medicinal chemists.

the SMILES representations of the selected molecules, obtained from PubChem, will be input into ProTox 3.0 to predict their toxicity profiles and classifications **[65][59]**.

- **PLIP:**

The Protein-Ligand Interaction Profiler (PLIP) is a sophisticated web-based tool designed to automatically identify and visualize non-covalent interactions between proteins and ligands within three-dimensional (3D) structures. It serves as an essential resource for researchers in structural bioinformatics, drug discovery, and molecular biology by providing comprehensive insights into protein-small molecule interactions **[59].**

F) Energy of Bonding and Antibonding Orbitals (HOMO and LUMO) :

In both physical and organic chemistry, understanding the electronic state of atoms is crucial for analyzing molecular interactions. The formation of chemical bonds results from atomic interactions, influencing molecular stability and reactivity. The study of these electronic states helps elucidate the mechanisms of synthetic reactions in organic chemistry.

Molecular bonding is described using two key electronic orbitals: the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO). These orbitals, also referred to as bonding and antibonding orbitals, play essential roles in chemical reactions. The HOMO is associated with bonding interactions, while the LUMO is involved in antibonding interactions, significantly influencing molecular reactivity and stability **[59].**

Conclusion :

Molecular docking has become a crucial tool in drug discovery, enabling researchers to predict how ligands interact with target proteins and optimize binding affinities. By combining computational techniques like Density Functional Theory (DFT) for molecular optimization with powerful software such as AutoDockVina, scientists can accurately simulate and refine drug-receptor interactions. Essential databases like PubChem and the Protein Data Bank (PDB) provide access to vast chemical and structural data, while tools like SwissADME, SwissTarget Prediction, and ProTox 3.0 help assess drug-like properties, biological targets, and toxicity risks. Additionally, analyzing molecular orbitals (HOMO and LUMO) offers deeper insights into reactivity and stability, further guiding drug design. Together, these methods enhance efficiency in virtual screening and structure-based drug development, reducing reliance on costly experimental trials. As computational approaches advance, molecular docking continues to drive innovation in pharmaceutical research, accelerating the discovery of safer and more effective therapeutics.

Chaptre III : Quantitative structure-activity relationship (QSAR)

Introduction:

The quantitative structure-activity relationship (QSAR) approach is grounded in a fundamental principle of chemistry: the biological activity of a molecule is intrinsically linked to its atomic arrangement and overall molecular structure. This concept posits that structurally similar compounds tend to exhibit comparable biological activities. Structural information is quantified using a set of parameters known as molecular descriptors. In QSAR modeling, biological activity is expressed as a function of these molecular descriptors, as illustrated in Equation 1:

The equation 1 **[**Biological activity = f(molecular descriptors)**]** represents the fundamental principle of Quantitative Structure-Activity Relationship (QSAR) modeling. It suggests that the biological activity of a chemical compound—such as its effectiveness as a drug or its toxicity—can be predicted as a mathematical function of its molecular descriptors. These descriptors are numerical values that encode information about the compound's physical, chemical, topological, or electronic properties. By analyzing these descriptors through statistical or machine learning models, researchers can identify patterns that relate molecular structure to biological effect, thereby enabling the design of more effective and safer pharmaceutical agents.

The resulting model, constructed from the biological activities of known compounds, can subsequently be employed to predict the biological responses of novel molecules.

A typical QSAR study involves several critical steps:

1. selection of compounds with known biological activity (both active and inactive),
2. calculation and analysis of molecular descriptors,
3. feature selection to identify the most relevant descriptors,
4. model construction using appropriate mathematical methods,
5. and thorough evaluation of the model’s predictive performance**[66].**

1. Quantitative structure-activity relationship (QSAR):

Quantitative structure-activity relationship (QSAR) analysis is a computational approach that establishes a correlation between chemical structures and their biological activities. The fundamental principle underlying QSAR is that variations in molecular structure result in

corresponding changes in bioactivity within a series of compounds **.**QSAR methods are widely employed in drug discovery to identify chemical entities with potential inhibitory effects on biological targets, while simultaneously minimizing toxicity, particularly in cases involving non-specific activity**[67]**.

The advancement of QSAR methodologies led to the development of three-dimensional QSAR (3D-QSAR) in the 1980s, incorporating 3D structural information to enhance predictive power. With the progress in computational technologies during the 1990s and the elucidation of high-resolution three-dimensional structures of biomacromolecules, structure-based drug design has increasingly supplanted traditional QSAR methods. Nevertheless, QSAR remains advantageous due to its relatively low computational demands and robust predictive capabilities**[67]**.

QSAR analysis begins with the determination of various molecular descriptors—quantitative representations of chemical structure—and the corresponding experimental biological activities. A statistical model is then developed to correlate molecular descriptors with observed bioactivity, allowing the prediction of biological properties for novel compounds Data Warrior, an open-source software developed by Open Molecules, facilitates structure-based data analysis and visualization. It computes a range of molecular descriptors, including molecular weight, lipophilicity, polar surface area, shape index, and molecular complexity, among others. DataWarrior is available for download at (https://openmolecules.org/datawarrior/download.html)**[67]**.

For the construction of a reliable QSAR model, several criteria must be satisfied:

(a) Bioactivity data should be available for at least 20 compounds, acquired through a standardized experimental protocol to ensure comparability of potency values;

(b) The division of compounds into training and test sets should be done judiciously;

(c) Molecular descriptors should exhibit minimal autocorrelation to avoid overfitting.

Model validation, essential to assess the applicability domain and predictive performance, can be achieved through internal or external validation methods**[67].**

The choice of statistical modeling approach depends on the nature of the data. Classification-based methods are applied to categorical (graded) data, whereas regression-based methods are employed for continuous quantitative data. Regression predicts continuous outcomes, while classification predicts discrete class labels. Among regression-based techniques, multiple linear regression (MLR) is the most widely used for developing QSAR models. MLR assumes a linear relationship between a dependent variable (biological activity) and multiple independent variables (descriptors). However, MLR can become computationally intensive when handling a large number of descriptors due to the stepwise selection process**[67]**.

Principal component analysis (PCA) offers a solution by reducing the dimensionality of the dataset, summarizing multiple descriptors into a smaller number of principal components. Nonetheless, the drawback of PCA is the reduced interpretability regarding which specific chemical features drive biological activity**.** Partial least squares (PLS) regression addresses the limitations of both MLR and PCA by projecting both descriptors and biological activities into new latent variables that maximize the correlation between them **[67].**

In bioinformatics, statistical modeling for QSAR studies is frequently performed using R software. Functions such as `lm` are commonly used for linear regression analysis, while PCA can be implemented using built-in functions like `prcomp` and `princomp`**.** Additionally, packages like dplyr facilitate data manipulation ,and the plspackage provides tools for PLS analysis**[67]**.

Validation is a critical phase in QSAR model development. The dataset used to build the model is termed the "training set," while a separate set, termed the "test set," is used to evaluate the model's predictive performance. Validation strategies include internal and external validation. Internal validation methods, such as leave-one-out cross-validation (LOO-CV), iteratively omit one molecule at a time from the training set to assess model robustness. Each omitted molecule’s activity is predicted using a model trained on the remaining compounds, and the process is repeated for all molecules. In contrast, external validation involves evaluating the model using an independent test set of compounds not involved in model construction, providing a more stringent assessment of predictive capability **[67]**

2. Classification of QSAR Approaches

QSAR methodologies can be categorized along three principal axes. The primary axis distinguishes methods based on the computational approach employed for **molecular descriptor derivation**. The second axis reflects the fundamental **chemical perspective** adopted within the study **[68].**

Table1 : classification and characterisation of the chemical approach**.**

Type of QSAR	**characteristic of the chemical approach.**
1D-QSAR	Uses unidimensional descriptors like pKa, aqueous solubility, logP, and functional group counts**[68].**
2D-QSAR	Based on topological indices such as connectivity and Wiener indices **[68]**.
3D-QSAR	Considers 3D spatial arrangement; focuses on steric and electrostatic interaction fields around the molecule**[68].**
4D-QSAR, 5D-QSAR, and 6D-QSAR	Incorporates multiple conformations, stereoisomers, tautomers, and solvation models for a more dynamic molecular representation**[68].**

Based on the categorization of chemical data analysis methodologies, QSAR methods are broadly classified into linear and non-linear approaches. The linear category encompasses techniques such as Multiple Linear Regression (MLR), Stepwise MLR (S-MLR). Conversely, the non-linear category includes methodologies such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) **[68].**

Finally, according to the third classification axis, a QSAR model can be categorized as either a **single-target QSAR model** or a **multi-target QSAR model**. Contemporary research in drug design increasingly emphasizes the **multi-target drug concept**, as highlighted in recent studies **[68, 69].**

3. Quantitative structure-activity relationship (QSAR) model construction

The predictive power of a QSAR model is significantly contingent upon the quality of the dataset employed for its development. Consequently, a comprehensive understanding of the data is a prerequisite prior to model construction. Thorough elucidation of the problem domain and influencing variables facilitates the identification of meaningful structure-activity relationships. Relevant background information pertaining to the biological or chemical system under investigation must be acquired through a comprehensive literature review. Careful selection of datasets for model construction is crucial, as suboptimal or inconsistent data will invariably lead to a compromised model. Furthermore, several other factors, including the partitioning of data into training and validation sets, the selection of appropriate molecular descriptors, and the statistical methodologies employed for model development, exert a substantial influence on the overall quality of the resulting QSAR model**[70]**.

A schematic representation of the QSAR model construction process is provided, and a concise description of the constituent steps involved in QSAR model generation is elaborated upon in the subsequent section **[70].**

3.1. Primary Data Collection

In practical terms, the development of QSAR models initiates with an initial phase focused on the availability and quality of experimental data, along with their respective quantities. This stage involves establishing a curated set of reference experimental data for the construction of a foundational database relevant to the property under investigation. Given the fundamental reliance of QSAR models on these reference experimental data, their selection is of paramount importance. The dataset must comprise data acquired through a standardized and well-documented experimental protocol. Consequently, the exclusion of low-quality or inconsistent data is essential to maximize the predictive power and robustness of the final model [**68].**

Furthermore, the following criteria are ideally met to ensure a sufficiently robust dataset:

(1) The range of values for the studied property should be maximized to define a broad and reliable domain of applicability for the resulting model. A wider range of property values within the training data enhances the model's potential to make predictions across a broader spectrum.

(2) It is generally advantageous for the distribution of the property values to approximate a normal distribution, as many standard statistical methods assume or perform optimally under this condition. However, it is important to note that certain non-linear methodologies can effectively handle non-normally distributed data, and the actual impact of the distribution depends on the specific statistical approach employed.

In many instances, experimental data are sourced from the peer-reviewed scientific literature. Numerous online databases aggregate information on a vast number of molecules. However, the construction of a high-quality dataset often necessitates a meticulous and systematic review of the original reference publications, coupled with the examination of diverse sources of structural information (such as crystallographic data, computational chemistry outputs, or SMILES representations), to ensure consistency and accuracy, rather than relying solely on Protein Data Bank (PDB) files, which primarily focus on the structures of biological macromolecules **[68]**.

3.2. Training and test data sets

The available dataset is strategically partitioned into independent training and test subsets. The training subset serves as the empirical basis for the derivation and parameterization of the QSAR model. Subsequently, the test subset is utilized for a rigorous, out-of-sample evaluation of the model's predictive capacity and accuracy. Typically, the data partitioning strategy aims to ensure a representative coverage of the entire descriptor space within both the training and test subsets. The judicious implementation of appropriate data splitting methodologies is critical for optimizing the model's ability to generalize to unseen data and ensuring the reliability of its predictions.

Various established techniques exist for dataset partitioning, including k-means clustering (based on the independent variable space, X), stratified sampling (based on the dependent variable, Y), random sampling, statistical molecular design principles, sphere exclusion algorithms, the Kennard-Stone algorithm, Kohonen's self-organizing map-based selection, and extrapolation-oriented test set selection strategies **[70, 71].**

3.3. Calculation of molecular descriptors

The structural information of molecules, encoded by molecular descriptors derived from various representations such as two-dimensional (2D) and three-dimensional (3D)

frameworks, is integral to the formulation of QSAR models. Accurate molecular conformations are crucial for developing robust predictive models.

A diverse array of molecular descriptors exists, categorized by their dimensionality and the type of molecular information they encode. These include count descriptors (0D), fingerprints (1D), topological descriptors (2D), geometrical descriptors (3D), and grid-based descriptors (4D), among others. Generally, the complexity of the encoded information and the capacity to differentiate between structurally similar compounds increase with the dimensionality of the descriptors. Zero-dimensional (0D) and one-dimensional (1D) descriptors provide fundamental information such as molecular weight and elemental composition, directly derived from the molecular formula. The net charge of a molecule is an example of a 1D descriptor.

Two-dimensional (2D) topological indices, computed from the molecular graph, are based on graph theory and reflect the connectivity patterns within the molecular structure. A widely utilized topological descriptor is the connectivity index proposed by Randić**[72**]. Other established topological indices include Wiener's index **[73],** Connectivity indices [**74**] , Kier Shape indices [**75**] , the Balaban J index [**76**] , and Zagreb indices [**77**] .

Three-dimensional (3D) descriptors are derived from the three-dimensional atomic coordinates of the constituent atoms. Common methodologies for calculating 3D descriptors include Comparative Molecular Field Analysis (CoMFA), Comparative Molecular Similarity Indices Analysis (CoMSIA), Comparative Binding Energy (CoMBINE), GRID-based Ensemble Receptor Mapping (GERM), Comparative Molecular Moment Analysis (CoMMA), Geometry, Representation, and k-based Index (GRIND), Weighted Holistic Invariant Molecular descriptors (WHIM), Hologram QSAR (HoloQSAR), and Comparative Spectra Analysis (CoSA).

Four-dimensional (4D), five-dimensional (5D), and six-dimensional (6D) descriptors are multidimensional, incorporating parameters related to the structure and flexibility of the receptor-binding site in conjunction with ligand topology. Four-dimensional descriptors are typically based on reference grids and molecular dynamics simulations. Five-dimensional descriptors are generated using multiple conformations, orientations, protonation states, and isosteres of the ligand. Solvation terms constitute six-dimensional descriptors. Numerous

software packages are available for the calculation of these diverse molecular descriptors, some of which are documented in the literature [**70**].

3.4. Classification of Descriptors

Molecular descriptors can be classified according to several criteria. A significant proportion of these descriptors are atom-based, as opposed to field-based. Atom-based descriptors are typically derived from the analysis of two-dimensional (2D) or three-dimensional (3D) connection tables and can encompass one-dimensional (1D), two-dimensional (2D), or three-dimensional (3D) information pertaining to the molecule. Categories of atom-based descriptors include individual atomic properties, counts of specific structural features, substructural fragments, topological indices, atomic properties, pharmacophoric features, and calculated physicochemical properties **[68]**.

Table2 :clasification and characterizations of discriptors

Type of discriptors	**Characterizations of discriptors**
Constitutionaldescriptors (1D)	Derived from the molecular formula; include counts of atoms (C, H, N, etc.), functional groups (NO_2, COOH, OH), bond types, rings, and weight.**[78, 68].**
2D descriptors	Calculated from molecular graphs; include topological indices (e.g., Wiener, Randić, Balaban) indicating volume, branching, and surface area. **[79, 82, 68]**
3D descriptors	Describe spatial atom arrangement; require molecular modeling using semi-empirical or ab initio methods **[68].**

These descriptors, while relatively computationally demanding, offer a richer representation of molecular characteristics and are essential for modeling properties or activities that exhibit a significant dependence on the three-dimensional (3D) molecular architecture. Severalkeyfamilies of 3D descriptorscanbeidentified:

- **Geometric descriptors:** Prominent examples include molecular volume, solvent-accessible surface area, and the principal moments of inertia. These parameters quantify the spatial extent and shape of the molecule.

- **Electronic descriptors:** A wide array of calculated physicochemical properties can serve as electronic descriptors, such as atomic electron densities, eigenvalues of frontier molecular orbitals (Highest Occupied Molecular Orbital - HOMO, and Lowest Unoccupied Molecular Orbital - LUMO), and Van der Waals volume and surface area. These descriptors reflect the electronic distribution and potential for intermolecular interactions.
- **Pharmacophores:** The identification of a pharmacophore is a critical step in elucidating the interaction mechanisms between a receptor and a ligand. According to the International Union of Pure and Applied Chemistry (IUPAC) definition (1998), a pharmacophore is "a set of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or block) its biological response" . In essence, pharmacophores are defined by the crucial physicochemical properties and potential or existing non-covalent interactions of molecules, rather than their complete structural formulas [**68**].

3.4.1.The Role of Molecular Descriptors in Malaria Drug Activity :

The fight against malaria, a persistent global health challenge, necessitates the continuous development of novel and effective therapeutic agents. A crucial aspect of rational drug design and optimization in this area lies in understanding the intricate relationship between the molecular characteristics of compounds and their ability to combat the *Plasmodium* parasite. This section delves into the significant role of molecular descriptors – quantifiable parameters that capture the structural, physicochemical, and electronic properties of molecules – in elucidating and predicting the activity of anti-malarial drugs. By exploring how these descriptors correlate with the potency of compounds against malaria, we aim to highlight their importance in guiding the development of more efficacious and selective treatments for this debilitating disease.

Table 3: Molecular Descriptors and Their Effect on Anti-malarial Activity

Type of discriptors	**Discriptors**	**The impact of descriptors on malaria**
	MolecularWeight – MW	Affects absorption, permeability, and drug kinetics.

2D	logP	Determines fat and water solubility, essential for membrane penetration and distribution.
	Wiener Index	Describe the molecule's shape and structure, influencing its biological interactions.
	Connectivity Indices	
	MolecularFingerprints	Represent the chemical structure and help identify key features for anti-malarial activity and the design of new drugs.
3D	Molecular Volume	Affects steric hindrance, target binding, and drug permeability.
	Solvent-Accessible Surface Area (SASA)	Determines hydrophilic and hydrophobic interactions and contributes to protein binding.
	Principal Moments of Inertia	Describe the molecule's shape and orientation, crucial for target binding.
	Radial Distribution Function – RDF	Defines the spatial atomic arrangement and interatomic distances, aiding in understanding interactions with parasite targets.
Quantum ChemicalDescriptors	HOMO /LUMO	Determining the ease of losing or gaining electrons, influencing chemical reactions and target binding.
	Polarizability	Defining the strength of van der Waals interactions and affecting binding and membrane crossing.
	Hyperpolarizability	Reflecting the molecule's response to electric fields and potentially contributing to specific interactions.
	Electrophilicity Index	Measuring the molecule's tendency to accept electrons, important for interactions with targets

		and avoiding toxicity.
Physicochemical Descriptors	Number of Hydrogen Bond Donors / Acceptors	Bind strongly and specifically to the parasite target, and achieve balanced solubility and permeability.
	Topological Polar Surface Area – TPSA	Permeate cell membranes and dissolve in water
	logS	Dissolve in bodily fluids and reach the parasite at an effective concentration.
	Refractive Index	Reflect molecular polarizability and density, and may correlate with activity in QSAR studies.

3.5. Feature selection

Feature selection is a dimensionality reduction technique that operates horizontally on the dataset, aiming to identify a subset of the most pertinent descriptors from the initially calculated pool for model development. This process is crucial for removing multicollinearity among descriptor pairs. The selection of the most informative features is typically accomplished using filter and wrapper methodologies **[83].**

Filter methods involve the independent evaluation of descriptors based on their intrinsic properties and relationships with the dependent variable, without involving a specific learning algorithm. These methods aim to reduce the descriptor pool size by filtering out less informative or redundant descriptors based on inter-variable correlations. Consequently, when descriptor pairs exhibit significant inter-correlation, only one representative descriptor is retained [**84**] . Descriptors exhibiting minimal variance across the dataset are also often eliminated. Common filter method techniques include chi-square analysis, Shannon entropy, odds ratio, the GSS coefficient [**85**], correlation-based feature selection [**86**] , Fisher Score, Kolmogorov-Smirnov statistics , and principal component analysis. Distance-based metrics, such as Euclidean distance measures, are also often categorized under filter methods.

Wrapper methods, in contrast, evaluate descriptor subsets by their performance within a specific predictive model. These methods employ regression-based or classification-based algorithms to iteratively select and evaluate different combinations of descriptors. While

generally demanding greater computational resources, wrapper methods often yield superior predictive performance compared to filter methods due to their direct optimization for the chosen modeling technique. Commonly employed wrapper methods include recursive feature elimination [**87**], variable selection and modeling based on prediction performance [**88**],

k-nearest neighbors, backward elimination, forward selection, genetic algorithms, Bayesian regularized neural networks, factor analysis, and combinatorial protocols.

Hybrid methods, which integrate aspects of both filter and wrapper approaches, are increasingly being utilized to leverage the efficiency of filter methods for initial dimensionality reduction and the enhanced predictive power of wrapper methods for refined feature subset selection **[70].**

4.Quantitative structure-activity relationship (QSAR) Methods

In QSAR modeling, statistical methods play a crucial role in both feature selection and model construction, particularly when dealing with large sets of molecular descriptors. These methods are instrumental in identifying functional relationships between chemical structure and biological endpoints. Broadly, statistical techniques used in QSAR can be classified into regression-based approaches, classification-based approaches, and machine learning methodologies. QSAR models can be developed to describe both linear and non-linear relationships between descriptors and biological activity.

For modeling linear relationships, techniques such as linear regression and partial least squares (PLS) regression are commonly employed. In contrast, non-linear relationships are often modeled using more advanced methods, such as artificial neural networks (ANNs), which are capable of capturing complex patterns within the data.

Model development can involve supervised, unsupervised, or semi-supervised learning techniques. Unsupervised learning methods tend to exhibit reduced risks of overfitting compared to supervised learning, as they do not adjust parameters to fit specific labeled data. Semi-supervised learning, which utilizes both labeled and unlabeled data, offers advantages over purely supervised or unsupervised methods by improving model performance and generalization ability**[68 83],[66].**

Comparative studies of supervised and semi-supervised learning algorithms across diverse datasets suggest that semi-supervised approaches can effectively leverage unlabeled data to enhance model performance. This is particularly beneficial for specific types of datasets and modeling tasks where labeled data are limited or expensive to obtain .**[69 84].**

4.1.Regression-Based Methods:

4.1.1.Multiple Linear Regression (MLR):

Multiple Linear Regression (MLR) is a statistical technique used to establish a correlation between independent variables (molecular descriptors) and a dependent variable (biological activity or physicochemical property) of the system under investigation. In simple linear regression, the prediction of the dependent variable is based on a single descriptor. However, in MLR, multiple descriptors are simultaneously considered to enhance the accuracy of the predicted property or activity.

The general form of a linear regression model can be expressed by the following equation:

$$\boldsymbol{y = a + bx} \qquad \textbf{(Equation 2)}$$

where:

- Y represents the dependent or response variable, corresponding to the biological activity or physicochemical property,
- x denotes the independent or predictor variable, representing the molecular descriptor, and
- b is the regression coefficient that quantifies the contribution of the descriptor to the response.

By incorporating multiple descriptors, MLR provides a more comprehensive model of the structure-activity relationship [70].

4.2. Support Vector Machines for Regression (SVR)

4.2.1 Support Vector Machines (SVM)

Support Vector Machines (SVMs), often termed Wide Margin Separators, constitute a class of supervised learning algorithms initially formulated for binary classification tasks, i.e., the prediction of a qualitative variable with two distinct categories. Subsequently, their

applicability was extended to quantitative variable prediction (regression). This methodology is rooted in the learning theory framework developed by Vapnik, commencing in 1995, with a central focus on controlling model complexity to optimize its generalization (predictive) capabilities [**85**]. SVMs gained rapid adoption due to their efficacy with high-dimensional datasets, their relatively small number of tunable parameters, their robust theoretical underpinnings, and their consistently strong empirical performance. Consequently, SVMs have been successfully applied across a diverse range of scientific and technological domains, including bioinformatics, information retrieval, computer vision, and finance [**68**].

4.2.2 SVM for Regression (SVR)

The fundamental principle of Support Vector Machine Regression (SVR) [**85**] parallels that of SVM classification: the objective is to minimize a regularized risk functional. The regularization term penalizes model complexity, typically inversely proportional to the square of the margin, while the loss is quantified using anϵ-insensitive function in the regression context.

This resulting function, f(x), aims to exhibit minimal deviation from the training instances (xi,yi), for i=1,…,N, while maintaining maximal flatness. Specifically, errors smaller than a predefined threshold ϵ are disregarded, and deviations exceeding ϵ are penalized [**86**].

Maximizing the flatness of the regression function serves to minimize the model's complexity, a critical factor influencing its ability to generalize to unseen data. Indeed, learning theory [85] posits that the generalization error can be bounded by the sum of two components: one directly related to the model's complexity and the other reflecting the error observed on the training dataset **[87, 88].** SVM methodologies inherently prioritize the control of model complexity during the training phase to enhance generalization performance [**68**].

5. Quantitative structure-activity relationship (QSAR) Model Validation

Validation constitutes a critical and indispensable aspect in establishing the reliability of QSAR models. Several distinct validation paradigms exist, encompassing internal and external validation methodologies. Contemporary research **[89]** underscores the necessity of internal validation as a fundamental step in model evaluation. Furthermore, external validation represents a crucial and requisite method for ascertaining both the generalizability

and the genuine predictive capacity of QSAR models [**90**]. More recently, Roy et al. **[91]** have proposed a set of parameters (QF12,QF22, QF32) as supplementary validation metrics.

5.1. Internal Validation

Beyond the root mean square error (RMSE) and the coefficient of determination (R2), the primary statistical parameters employed in this study to assess model performance include:

- The cross-validation coefficient (Qappr2 or QLOO2), which should ideally exceed 0.5.
- The metrics rm^2 and Δrm^2 as proposed by Roy et al. Established guidelines suggest that Δrm^2 should preferably be less than 0.2, provided that rm^2 is greater than 0.5 [**68**].

5.2. External Validation

The most commonly employed external validation parameters and criteria are outlined below:

- **RMSE and R^2:** These standard metrics assess the magnitude of prediction errors and the proportion of variance explained by the model in the external test set.
- **The $Qtest^2$(or $Qext^2$) coefficient:** This parameter quantifies the correlation between the observed and predicted activity data for the external test set. Models exhibiting Qtest2 values exceeding a threshold of 0.5 are generally considered to possess acceptable predictive power.
- **Golbraikh and Tropshacriteria :** These authors proposed a set of stringent criteria to ascertain the external predictability of QSAR models. According to these guidelines, a model is deemed satisfactory if all the following conditions are met:

- $R^2 > 0.6$
- $Qtest^2 > 0.6$
- $(R^2 - Qtest^2)/R^2 > 0.1$
- $R0^2 - R0'^2 < 0.3$
- $0.85 \leq k \leq 1.15$or$0.85 \leq k' \leq 1.15$ where $R0^2$ and $R0'^2$ are the squared correlation coefficients through the origin (for the predicted vs. observed and observed vs. predicted plots, respectively), and k and k′ are the slopes of these regression lines.

- **The Concordance Correlation Coefficient (CCC):** This parameter serves to evaluate the reliability of the model by quantifying both the dispersion of the observations around the line of perfect agreement and the deviation of the regression line from the line passing through the

origin (the concordance line). Any deviation of the regression line from the concordance line results in a CCC value less than 1 **[92, 93] [68].**

The equations for these different internal and external validation parameters **[68]** and the meaning of the various terms are given below :

$$RMSE = \sqrt{\frac{\sum(Yexp-Ypred))^2}{n}}$$ (Equation 3)

$$R^2 = 1 - \frac{\sum(Yexp-Ypred)^2}{\sum(Yexp-\overline{Y})^2}$$ (Equation 4)

$$Qapp^2 = 1 - \frac{\sum(Yexp(appr)-Ypred(test))^2}{\sum(Yexp(test)-\overline{Y\,\exp(appr)})^2}$$ (Equation 5)

$$Qtest^2 = 1 - \frac{\sum(Yexp(test)-Ypred(test))^2}{\sum(Yexp(test)-\overline{Y\exp(appr)})^2}$$ (Equation 6)

$$r_{m^2=r^2(1-\sqrt{r^2-r_0{}^2}}$$ (Equation 7)

$$r'_{m^2=r^2(1-\sqrt{r^2-r'_0{}^2}}$$ (Equation 8)

$$\overline{r_{m^2}} = \frac{(r_{m^2}+r_{m^2})}{2}$$ (Equation 9)

$$\Delta r_{m^2} = |r_{m^2}-r'_{m^2}|$$ (Equation 10)

$$CCC = \frac{2\sum(Yexp(test)-\bar{Y}\exp(\text{test}))(\text{Ypred(test)}-\bar{Y}\text{pred(test)}}{\sum(Yexp(test)-\bar{Y}\exp(\text{test}))^2+\sum(Ypred(test)-\bar{Y}\text{pred(test)}))^2}$$(Equation 11)

Where Yappr andYtest represent the experimentally determined values of the studied property for the training and test sets, respectively, and Y^appr and Y^test denote the corresponding predicted values. The variable n signifies the number of compounds (or observations) within the respective dataset (training or test). Finally, $\bar{Y}$ represents the mean value of the studied property within the specified dataset (training or test) and for the designated data type (experimental or predicted).

Conclusion :

The Quantitative Structure-Activity Relationship (QSAR) approach stands as a robust and computationally efficient methodology in drug discovery and chemical research, establishing correlations between molecular structure and biological activity via quantitative descriptors. The QSAR paradigm encompasses a sequence of critical steps: the selection of compounds with experimentally determined activities, the calculation of relevant molecular descriptors, feature selection to identify the most predictive variables, the construction of a statistical or machine learning model, and a thorough validation process to ascertain the model's reliability and predictive power. QSAR models span a spectrum of dimensionality, from one-dimensional (1D) to higher-dimensional techniques such as three-dimensional QSAR (3D-QSAR), which integrate spatial and electronic properties to enhance predictive accuracy for complex biological interactions. Common modeling algorithms include Multiple Linear Regression (MLR) and Support Vector Machines (SVM), while rigorous validation protocols, employing both internal (e.g., cross-validation) and external (e.g., independent test set evaluation) metrics, are essential to ensure the model's robustness and generalizability. Despite significant progress in structure-based drug design, QSAR continues to be a valuable tool due to its computational expediency and robust predictive capabilities, rendering it an indispensable asset in contemporary pharmaceutical research.

Chapter IV: Results and Discussion - Computational and Experimental Evaluation of Plant Compounds as Inhibitors of Malaria target Proteins

I).Methodological Overview of in silico molecular docking' process:

Molecular docking is a pivotal computational technique in modern drug discovery, enabling researchers to predict the interaction between small molecules (ligands) and target proteins at the atomic level. By simulating the binding process in three-dimensional space, molecular docking provides insights into the structural basis of molecular recognition, binding affinity, and potential therapeutic efficacy of compounds. This section explores the principles of molecular docking, the role of optimization techniques like Density Functional Theory (DFT), and the application of tools such as AutoDockVina and CB-Dock to evaluate interactions between bioactive compounds (e.g., Artemisinin, Chamazulene, Thymol, and Carvacrol) and target proteins (e.g., PF-NDH2, plasmepsins). Additionally, it examines physicochemical properties, toxicity profiles, and bioactivity predictions using databases like SwissADME and ProTox 3.0, offering a comprehensive framework for identifying promising drug candidates.

Bioinformatics Tools and Databases:

a)-Pubchem:

Four compounds were selected: Artemisinin from *Artemisia annua*, Chamazulene from *Artemisia afra*, and Thymol and Carvacrol) from *Thymus vulgaris*.These compounds were downloaded from a drug database called PubChem form SDF, An open-access chemistry database and a drug bank comprising substances, compounds, and bioassays. (https://pubchem.ncbi.nlm.nih.gov/)The following compounds were downloaded:

Artemisinin-CID: 68827 ;chamazulene- CID: 10719 ; thymol-
CID: 6989 ;carvacrolCID: 10364

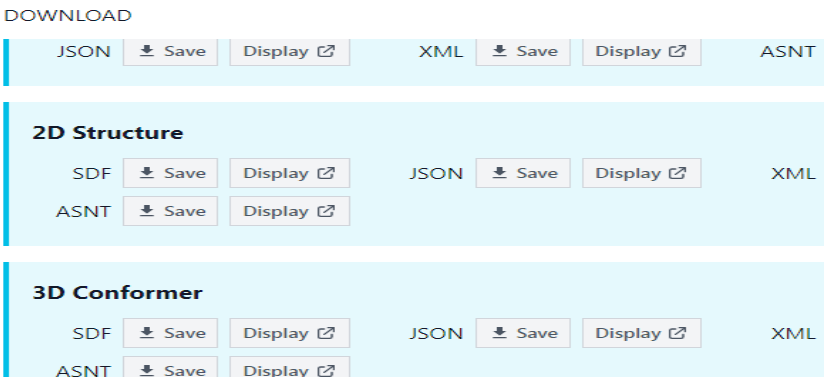
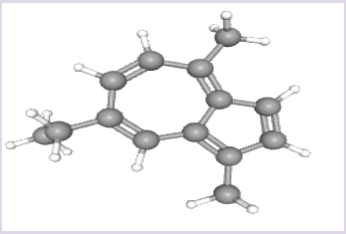
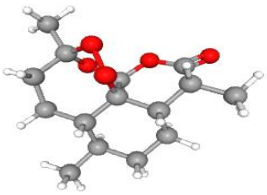
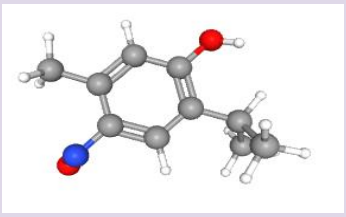
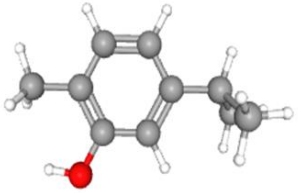


Figure 14: download molecule format SDF 3D

Table4: molecules obtained from PubChem

Plant	Ligand	SDF Form
Artimisiaafra	Chamazulene	
Artemisiaannua	Artemisinin	
Thymus vulgaris	Thymol	
	Carvacrol	

b) Preparation of ligand :

- **Optimization (DFT) :**in this step, molecular structures will be input into Gaussien software for optimization.theoptimisation process follows the steps illustrated (in figure

24).Structural optimizationwas done using Spartan'14 software with DFT/B3LYP at 6-311G* levels. This was done to obtain the frontier orbital energies, the highest occupied molecular orbital (HOMO) the lowest occupied molecular orbital (LUMO) and energy gap (in figure 25). Information on the stability and reactivity of the compounds are also provided **[60].**

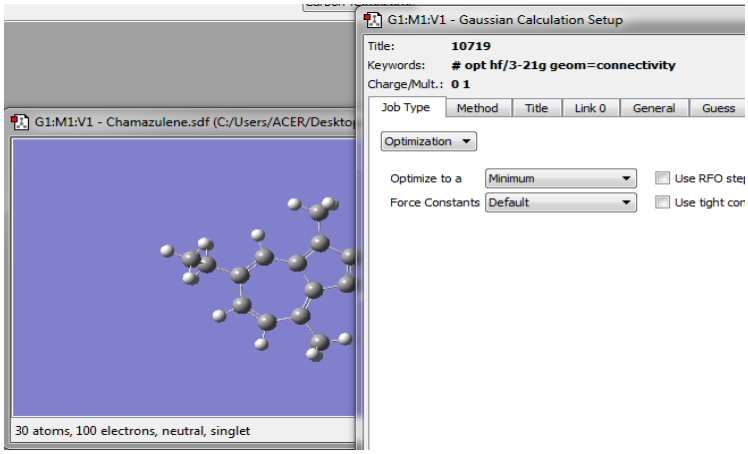
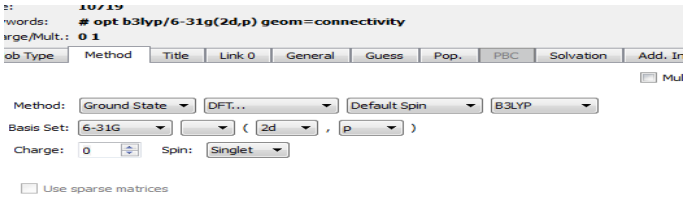
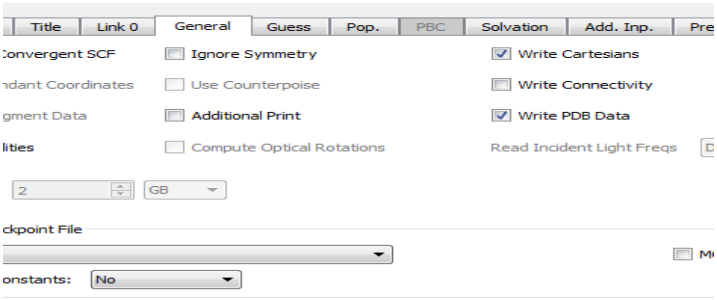


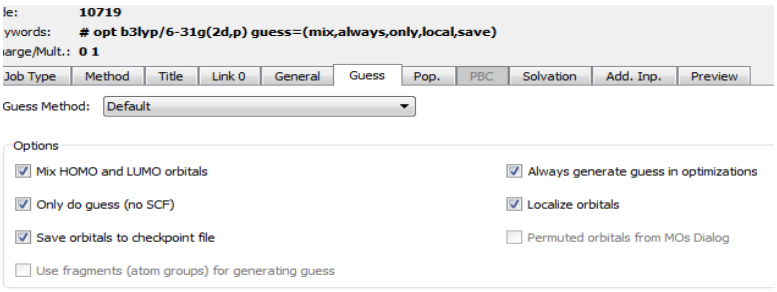
Figure15: Optimization step with Gaussien.



(a)



(b)



(c)

Figure16: (a) DFT method (b) unhook write connectivity (c) click the 5 lockers

After initiating the optimization calculation process (see figure 26),we wait for the computation to complete

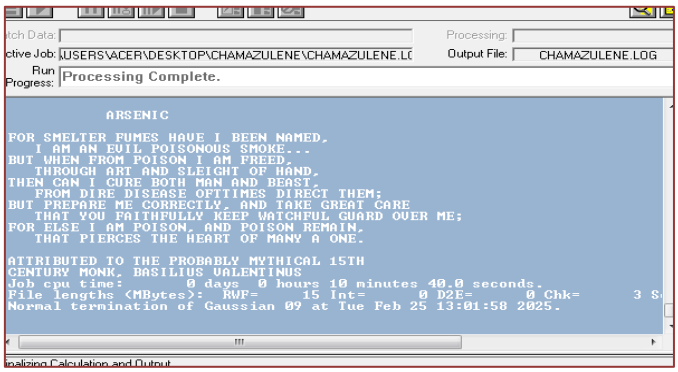


Figure 17: calculation period for optimization

Theresults are then generated in the GJF format,which is subsequently converted into PDB format(see figure 27).

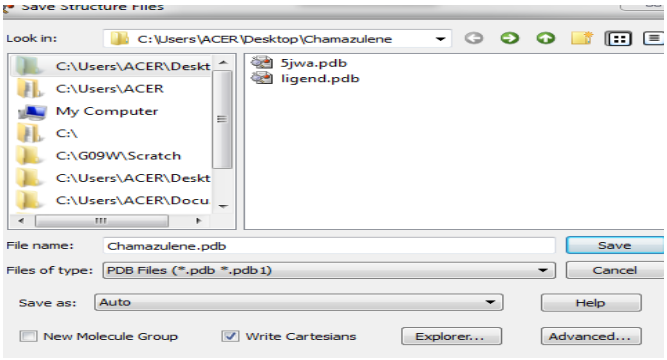


Figure 18: Form GJF converts to PDB

c) Preparation of target:

in this step,we will download (https://www.rcsb.org/) (see figure28) the target protiens or receptors (PF-NDH2, plasmepsins, PF-plasmepsin_2, and Apicoplast DNA Polymerase).

- PF-NDH2: **5jwa**
- Plasmepsins: **2IGX**
- PF-plasmepsin_2: **2IGY**
- Apicoplast DNA Polymerase: **7QXS**

Figure19:download the targets (pdb format)

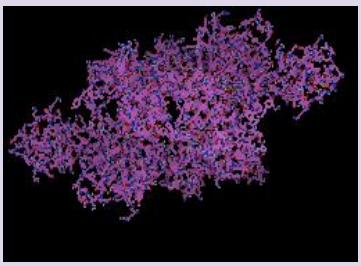
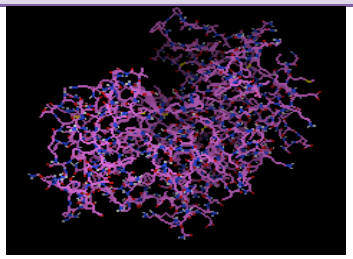
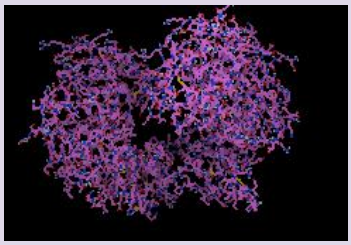
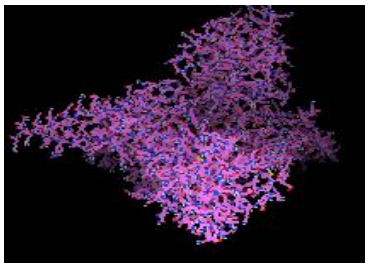
e) In silico (molecular docking)

1- Presentation of protein and ligand:

Receptors (The proteins/target): the protiens utilized in this study include protien

PF-NDH2, plasmepsins, PF-plasmepsin_2, and Apicoplast DNA Polymerase ,which were obtained from the protein data bank (PDB) in their three –dimentional (3D) structures. These proteins were prepared according to the previously described procedures and converted into the PDBQT format for molecular docking analysis. Table 05 presents the processed target proteins in PDBQT format, prepared using the AutoDockVina tool.

Table5: target pdbqt format by autoDockVina tool

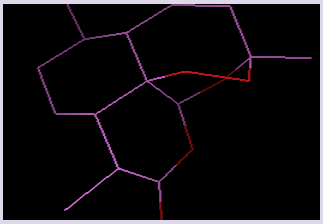
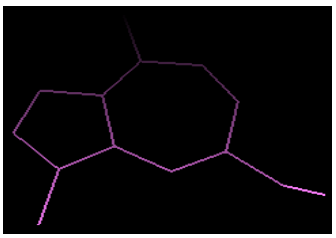
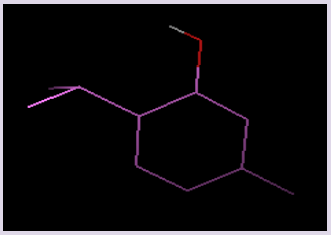
Target	Format PDBQT
PF-NDH2	
Plasmepsins	
PF-plasmepsin_2	
Apicoplast DNA Polymerase	

f) Ligand preparation:

We prepared the compounds ***(*Artemisinin *from Artemisia annua,* Chamazulene *from Artemisia afra, and* Thymol and Carvacrol *from Thymus vulgaris.)***

In Guaussian / Gaussview and Autodock tools by converting them from pdb-out to pdbqt as previously described, and got the results (in the table 6).

Table 6: molecular plants in AutoDockVina tool

Plants	Molecule	Format PDBQT
Artemisia annua	Artemisinin	
Artemisia afra	Chamazulene	
Thymus vulgaris	Thymol	
	Carvacrol	

g) InSilico interaction between ligand pdbqt& target pdbqt

Molecular docking studies revealed promising binding interactions between bioactive plant compounds and essential Plasmodium proteins, unveiling novel avenues for antimalarial drug development. Precise binding sites were identified within catalytic domains of target proteins, suggesting selective inhibition of their biological functions:

- **Chamazulene** interacts with site (1.763, -47.973, 7.397) in the **NADH binding region** of *Plasmodium falciparum***PF-NDH2**, suggesting a potential disruption of the parasite's **mitochondrial electron transport chain**. This is consistent with previously reported **sesquiterpene inhibition of redox enzymes**.
- Alternatively, **thymol** binds at (65.311, 56.412, 6.704), near the **catalytic aspartate dyad (Asp34/Asp214)** of **plasmepsins**—key proteases in the parasite's life cycle. This supports **competitive active site inhibition** and aligns with the known ability of **phenolic compounds** to form hydrogen bonds with proteases.
- For **Artemisinin**, a binding location (53.925, -0.358, -18.487) is observed overlapping with the flap region of PF-plasmepsin_2, potentially hindering the movement of this structural element necessary for substrate binding, and providing a mechanistic insight The findings highlight the significance of **the observed synergism between artemisinin and protease inhibitors**, suggesting a promising strategy for enhanced antimalaria.
- Finally, **carvacrol**, a **phenolic monoterpenoid compound**, displays a binding site (-49.129, 3.346, 18.507) within the **nucleotide binding cleft** of *Apicoplast DNA Polymerase*, a crucial enzyme in the parasite's apicoplast. This suggests its potential interference with the polymerization mechanism via its ability to form **hydrogen bonds** and engage in **hydrophobic interactions**, which is consistent with the known polymerase inhibitory properties of terpenes.In this step, we will perform molecular docking and extract the grid box (see table 7)

Table 7: molecular docking (InSilico) and show grid box

Target PDBQT	Ligand PDBQT	Grid box
PF-NDH2	Chamazulene	center_x = 1.763 center_y = -47.973 center_z = 7.397
Plasmepsins	Thymol	center_x = 65.311 center_y = 56.412 center_z = 6.704
PF-plasmepsin_2	Artemisinin	center_x = 53.925 center_y = -0.358 center_z = -18.487
Apycoplast DNA Polymerase	Carvacrol	center_x = -49.129 center_y = 3.346 center_z = 18.507

h) AutoDockVina:

Overall, these Vina scores support the proposed mechanistic interpretations based on the binding sites. Stronger binding (more negative Vina scores) indicates a higher likelihood of a relevant biological interaction. However, it is important to note that molecular docking results are predictions and require further experimental validation to confirm these interactions and determine their pharmacological significance.

- **PF-NDH2 + Chamazulene (-7.7 kcal/mol):** The most negative binding energy score here indicates a strong affinity between chamazulene and the PF-NDH2 enzyme. This suggests that chamazulene has a high propensity to bind to the previously mentioned NADH

cofactor region, thereby enhancing the likelihood of disrupting the enzyme's function and consequently affecting the parasite's mitochondrial electron transport chain.

- **PF-plasmepsin_2 + Artemisinin (-7.4 kcal/mol):** This also shows a negative binding score, indicating a favorable interaction between artemisinin and the PF-plasmepsin_2 enzyme. This supports the potential for artemisinin to impact the flap region of this enzyme, as previously suggested, which could hinder substrate binding and contribute to its anti-parasitic effect.
- **Plasmepsine + Thymol (-7.2 kcal/mol):** This binding energy value suggests a notable interaction between thymol and the Plasmepsine enzyme. This supports the idea that thymol can bind to the enzyme's active site, as proposed based on the binding position, potentially leading to competitive inhibition.
- **Apicoplast DNA Polymerase + Carvacrol (-6.5 kcal/mol):** While still a negative binding energy, this score indicates a comparatively weaker affinity of carvacrol for the Apicoplast DNA Polymerase enzyme compared to the other compounds and their targets. Nevertheless, this binding could still be sufficient to suggest a potential interference with the nucleotide binding site and thus impact polymerase activity.

The Vina docking score results suggest varying binding affinities between the studied compounds and their parasitic protein targets (in table 08).

Table 08:vina score of interaction between ligand pdbqt and targetpdbqt

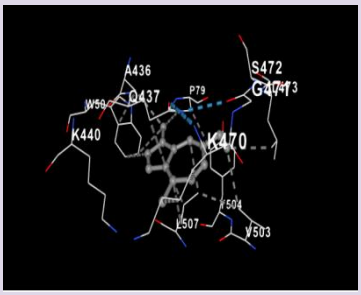
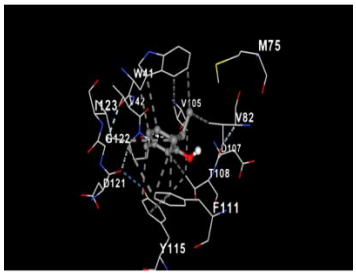
Ligand PDBQT &Target PDBQT	Vina score (kcal/mol)
PF-NDH2 + chamazulene	-7.7
Plasmepsine + thymol	-7.2
Apycoplast DNA Polyemerase +carvacrol	-6.5
PF-plasmepsin _2 +Artemisinin	-7.4

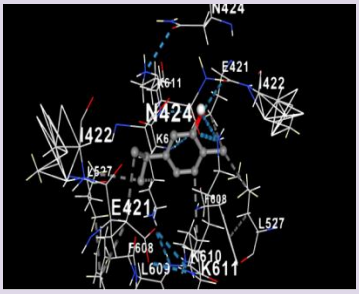
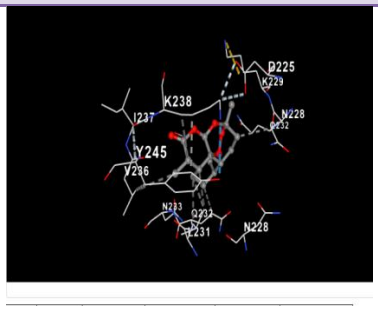
i) Databases:

- **CB-Dock:**

Table 09 represent the Molecular docking of elucidated the specific binding modes of four plant-derived compounds against crucial malaria parasite targets. Chamazulene, from *Artemisia*, interacts with the PF-NDH2 enzyme's NADH-binding site via π-π stacking (W50, Y504), **hydrophobic contacts (V503, L473)** and ionic bonds (K440, K470), potentially disrupting the electron transport chain. While data is limited, thymol is predicted to inhibit plasmepsins' catalytic sites (Asp34/Asp214) through hydrogen bonds and hydrophobic interactions, consistent with its protease inhibitory activity. Artemisinin, beyond its known mechanism, directly binds to PF-plasmepsin_2's catalytic residue D225, forming salt bridges (K229, K238) and hydrophobic/polar interactions (L234, I222, N233, Q232). Carvacrol exhibits a unique dual binding pattern with the Apycoplast DNA Polymerase, targeting both the active site (D107, W41) and a regulatory region (E421, K610) with extensive **hydrophobic interactions (F608, L527)**, potentially offering a strategy against drug resistance.

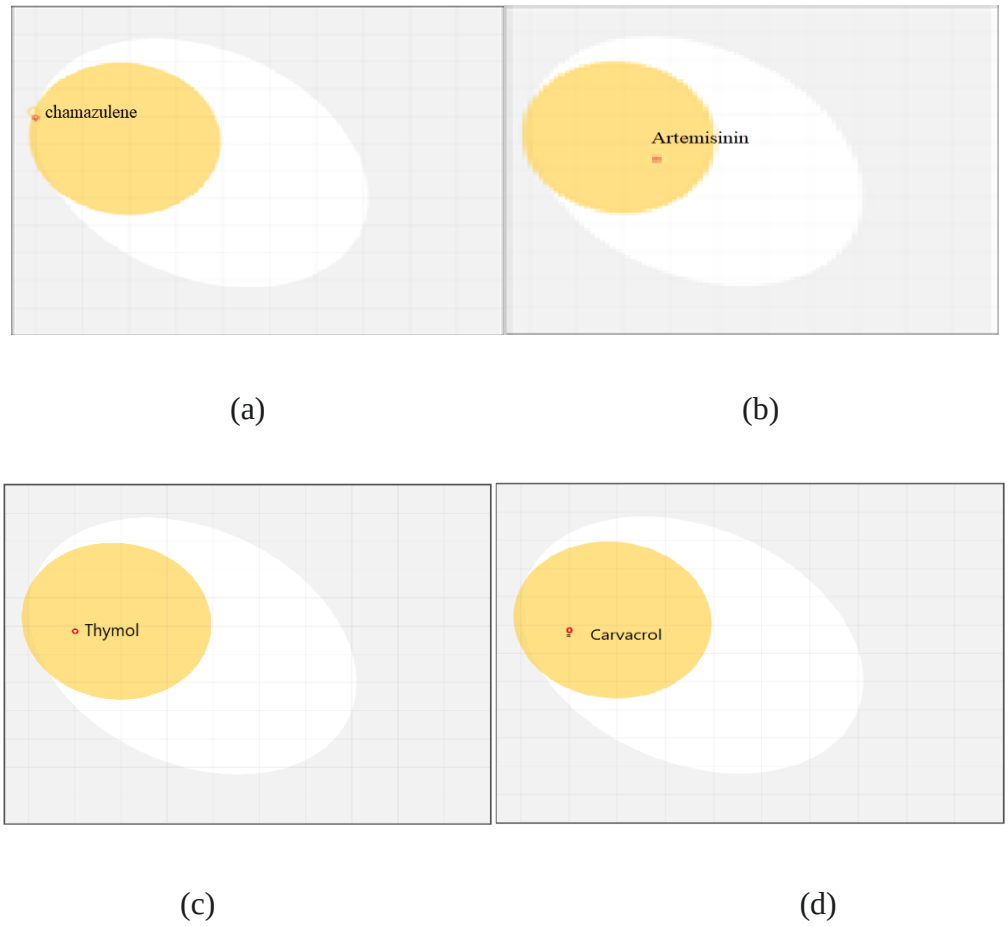
Table 09 : interactions usingCBdock**.**

Ligand &target	CB Dock interaction	Targetedamino-acids
PF-NDH2 + chamazulene		L507 ; Y504 ; V503 ; W50 ;K440 ; A436 ; L473 ; Q473 ; S472 ; K470 ; G471 ; P79
Plasmepsine + thymol		M75 ;W41 ; V105 ;V82 ; D107 ; T108 ;F111 ; Y115 ; D121 ; V42 ;W41 ;G122 ;M23 ; T45

Apycoplast DNA Polyemerase +carvacrol		E421 ;K611 ;K610 ;N424 ;F608 ;L527 ;K610 ;K611 ;E421 ;N424 ;L609 ;I422 ; I422 ;F608 ;L527
PF-plasmepsin _2 +Artemisinin		L234 ; N233 ; Q232 ; K229 ; K238 ; D225 ; N228 ; N228 ; K229 ; L231 ; V236 ; I222 ; Y245

- **SwissADME:**

In silico ADME predictions reveal distinct pharmacokinetic profiles for Thymol, Artemisinin, Carvacrol, and Chamazulene. Thymol and Carvacrol exhibit moderate lipophilicity and low Topological Polar Surface Area **(TPSA)**suggesting good passive membrane permeability but potentially limited by predicted low aqueous solubility and moderate bioavailability. Artemisinin shows a similar lipophilicity but a slightly higher TPSA and better predicted aqueous solubility. Notably, Chamazulene displays high lipophilicity and zero TPSA, indicating potentially excellent membrane permeability, including the skin and possibly the blood-brain barrier, although its aqueous solubility might be a concern despite a moderate predicted value. All four compounds have a predicted moderate oral bioavailability score. These computational insights highlight the need for experimental validation to confirm these ADME properties and further assess their potential as drug candidates, considering factors beyond these predictions (see table 10)



(a) (b)

(c) (d)

Figure20: diagram of boiled egg shows in Swissadme of our molecules.(a) Chamazulene, (b)Artemisinin, (c) Thymol, (d) Carvacrol

Table10: The ADME profiles vary in terms of lipophilicity and skin permeability

Ligand (molecule)	**ADME**	
Thymol	Lipophil	Log p=2.80 Log kp=-4.87 TPSA=20.23A°

Artemisinin	Lipophil	Log p=2.49 Log kp=-5.69 TPSA=53.99A°
Carvacrol	Lipophil	Log p=2.82 Logkp=-4.74 TPSA=20.23A°
Chamazulene	Lipophil	Log p=4.24 Log kp=-4.28 TPSA=0.00A°

- **Swiss Target Prediction:**

In this step, the SMILES representations of the selected molecules, obtained from PubChem, are submitted to SwissTargetPrediction to evaluate the likelihood of interactions between the molecules and their potential biological targets (see figure 30).

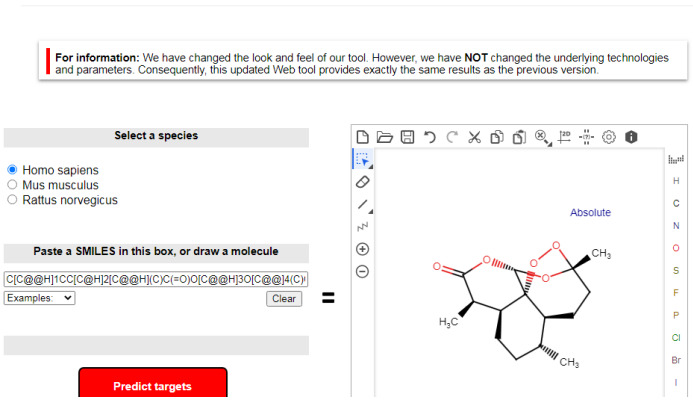


Figure 21:swiss terget prediction.

- **ProTox 3.0**

In this step, the SMILES representations of the selected molecules, obtained from PubChem, were input into **ProTox-T (ProTox 3.0)** to predict their toxicity profiles and classifications. The acute toxicity of **Thymol, Carvacrol, Artemisinin, and**

Chamazulene, as presented in Table 11, was assessed through their predicted **LD50 values** and corresponding **Toxicity (Tox) classifications** based on the Globally Harmonized System (GHS).

Among the tested compounds, **Artemisinin** exhibited the lowest acute toxicity, falling into **GHS Class 5** (Tox=05) with an LD50 of 4228 mg/kg. This indicates that a substantially higher dose is required to cause lethality in test animals compared to the other molecules. **Thymol** (Tox=04, LD50=640 mg/kg), **Carvacrol** (Tox=04, LD50=810 mg/kg), and **Chamazulene** (Tox=04, LD50=1220 mg/kg) all belong to **GHS Class 4**, indicating moderate acute toxicity. Although Chamazulene's LD50 value is higher than Thymol's within this dataset, all three are classified similarly as harmful if swallowed(see table 11).

Table11:predictLD50 and a toxicity category of molecules

Molecules	Classement of toxicity	LD50
Thymol	Tox =04	LD50= 640mg/kg
Carvacrol	Tox=04	LD=810 mg/kg
Artemisinin	Tox=05	LD=4228 mg/kg
Chamazulene	Tox=04	LD=1220 mg/kg

- **PLIP:**

In this step, the target protein structure in PDB format will be uploaded to PLIP, where the interactions will be analyzed, and the spatial positions and distances between the target and the ligand will be visualized.

The study of PLIP enables the indentification of optimal ligand binding sites on the target protien. It provides insights into hydrophobic interactions,which are considered the most favorable regions for ligand accumulation. For furtherdetails, refer to figure (31,32,33, 34).

❖ **PF-NDH2 :**

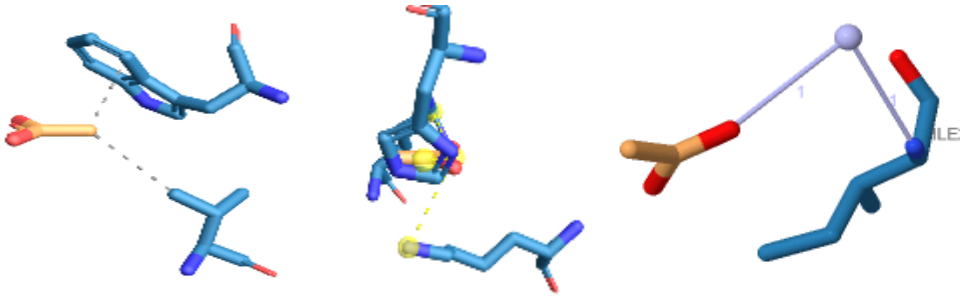


Figure 22: Moleculare interaction of the compounds selected from PF-NDH2

❖ **Plasmepsine :**

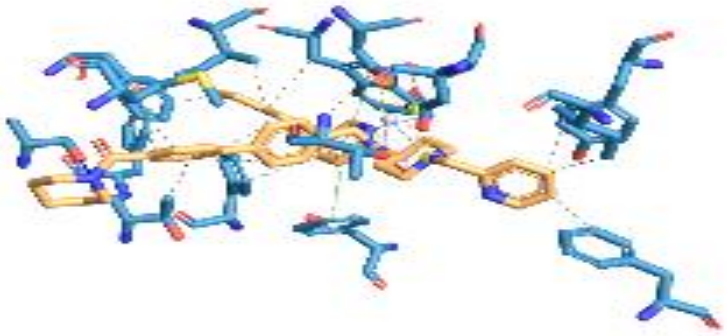


Figure 23: Moleculare interaction of the compounds selected from PLasmepsine

❖ **PF-plasmepsin_2 :**

Figure 24: Moleculare interaction of the compounds selected of PF-plasmepsin_2

❖ **Apycoplast DNA Polymerase :**

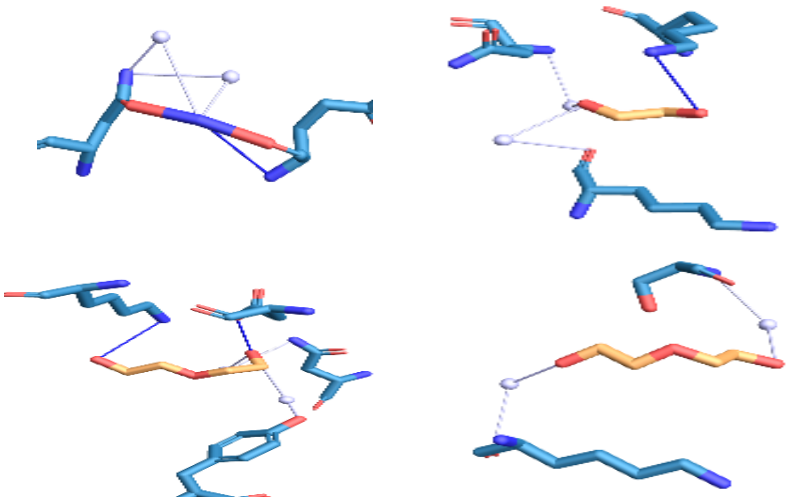


Figure 25 :Molculare interaction of the compounds selected from Apycoplast DNA Polymerase.

j) Lipinski's rule:

This section introduces Lipinski's Rule of Five, a set of guidelines used in drug discovery to predict the oral bioavailability of a molecule based on its physicochemical properties: having no more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors, a molecular mass under 500 Daltons, and a Log P not greater than 5. The provided table then lists these properties for Artemisinin, Carvacrol, Chamazulene, and Thymol, setting the stage for an evaluation of their potential oral bioavailability according to Lipinski's Rule (see table12).

We apply Lipinski’s Rule of five to our four molecule:

Table 12: Applying Lipinski of molecules

Molecule	Molecularweight	Log p	HA	HD
Artemisinin	282.33	2.49	5	0
Carvacrol	150.22	2.82	1	1
Chamazulene	184.28	3.97	0	0
Thymol	150.22	2.80	1	1

k) Energy of Bonding and Antibonding Orbitals (HOMO and LUMO):

This section explains the significance of HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital) energies in understanding molecular reactivity. The HOMO represents the electron-donating ability, while the LUMO represents the electron-accepting ability. The energy gap between them (LUMO - HOMO) indicates kinetic stability. The provided table shows these energy values for Artemisinin, Thymol, Chamazulene, and Carvacrol, allowing for a comparison of their relative tendencies to participate in chemical reactions based on their electronic structures. Smaller energy gaps generally suggest higher reactivity.

differences (**energy gaps**), which are crucial theoretical indicators for predicting the chemical behavior of the studied molecules. For instance, the lower energy gaps observed in **Carvacrol**and **Thymol** may suggest increased **reactivity** compared to **Artemisinin**. Conversely, **Chamazulene's** larger energy gap indicates higher chemical stability.

Table 13: Results of Energy of Bonding and Antibonding Orbitals (HOMO and LUMO):

Molecule	**HOMO**	**LUMO**	**ENERGIE OF GAP** **DE=LUMO – HOMO**
Artemisinin	-0.29331	-0.11397	0.17941
Thymol	-0.31382	-0.15286	0.16096
Chamazulene	-0.42259	-0.20330	0.21929
Carvacrol	-0.30791	-0.15139	0.15652

l) Bioactivity of Selected Compounds:

Table 14 presents the bioactivity parameters of the four selected hit compounds. An inverse relationship is observed between binding energy and inhibition constant (as described by Equation 12), where higher binding energy corresponds to a lower inhibition constant. This suggests that a potential hit compound should exhibit an inhibition constant within the range of 0.1–1.0 µM and should not exceed 10 nM for a highly potent drug**[94].**

Ki=e[-ΔG /RT] (Equation 12)

Where R=Gaz constant (1.987 *10-3 kcal/k-mol);

T=298.15(Absolute Temperature) ; Ki=Inhibition constant

The **inhibition constant (Ki)** quantifies the **binding affinity** between an inhibitor and its target enzyme or receptor. A **lower Ki value** signifies a **higher binding affinity**, meaning a smaller concentration of the inhibitor is required to effectively inhibit the enzyme.

In the context of drug potency, inhibitors with **Ki values below 100 µM** are generally considered **potent**, whereas those with **Ki values exceeding 100 µM** are typically classified as **non-potent**.

Analyzing the values of the compounds:(Artemisinin: M (3.76 µM), Carvacrol: M (3.37 µM), Chamazulene: M (2.27 µM), Thymol: M (7.39 µM)).

All these compounds have values well below 100 µM, classifying them as potent inhibitors. Among them, chamazulene exhibits the lowest value, suggesting it has the highest binding affinity and, potentially, the greatest inhibitory potency (see table 14).

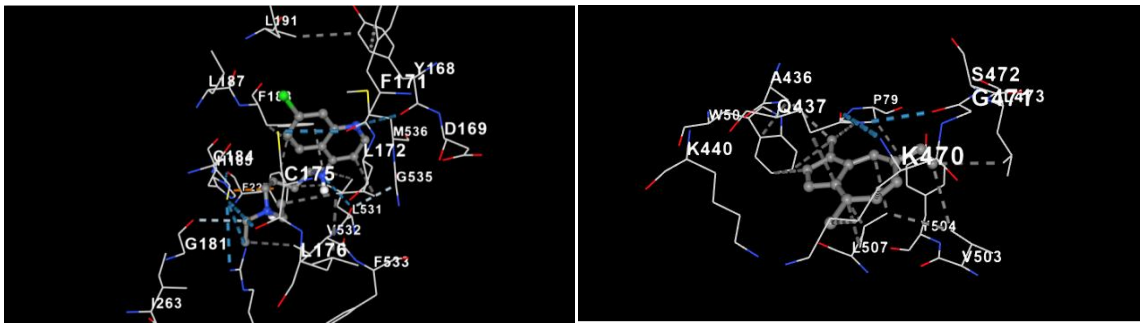
Table 14:inhibition constant of four compounds

COMPOUNDS	**CID**	**Binding Affinity (ΔG), kcal/mol**	**KI(µM)**
Artemisinin	68827	-7.4	3.76
Carvacrol	10364	-6.1	3.374
Chamazulene	10719	-7.7	2.26
Thymol	6989	-7	7.386

M) Comparison between the commercial treatment and our molecule:

1)antimalarialtreatementchloroquine and chamazulene :

- We compared the drug chloroquine using CBdock



(1) (2)

Figure 26: CB dock Interaction of (1) chloroquine and (2) chamazulen

Table 15: Comparison of Amino Acid Chains Interacting with Chloroquine and Chamazulene via CB-Dock Analysis

Amino acid chain of CB dock interaction of chloroquine :	**Amino acid chain of CB dock interaction of chamazulene :**
:L191 ;F171 ;Y168 ;F188 ; L187 ; G535 ; L531 ; C175 ;G184 ;F227 ; G181 ;V532 ; L531 ; L176.	L507 ; Y504 ; V503 ; W50 ;K440 ; A436 ; L473 ; Q473 ; S472 ; K470 ; G471 ; P79

2) bioactivity parameters (ki) of the chloroquine:

Binding Affinity (ΔG)=-8 kcal/mol

$$Ki=e[-\Delta G /RT] \quad \text{(Equation 12)}$$

Where R=Gaz constant (1.987 *10-3 kcal/k-mol);

T=298.15(Absolute Temperature) ; Ki=Inhibition constant

The constant inhibition of chloroquine:

$$Ki=1.3\mu M \quad \text{(Equation 13)}$$

- In comparison, while the studied molecules show promising in vitro properties, chloroquine remains a well-established and clinically effective drug against malaria, particularly before the emergence of widespread resistan

Summary of Results:

This study explores the potential of four plant-derived compounds (artemisinin, chamazulene, thymol, carvacrol) as antimalarial drugs by analyzing their interactions with essential proteins of the *Plasmodium falciparum* parasite using molecular docking techniques and in silico calculations. The results demonstrated promising binding interactions of these compounds with specific sites in the target proteins (PF-NDH2, plasmepsins, PF-plasmepsin_2, apicoplast DNA polymerase), exhibiting strong binding affinity and negative free energies. Furthermore, the pharmacokinetic properties (ADME) and potential toxicity of these compounds were evaluated using specialized databases, indicating varying pharmacokinetic profiles and acute toxicity ranging from low to moderate. Moreover, all compounds complied with Lipinski's Rule of Five for predicting oral bioavailability, and they showed low inhibition constant (Ki) values, classifying them as potent inhibitors. In comparison to the known antimalarial drug chloroquine, the studied compounds exhibited promising in vitro properties, warranting further experimental validation to assess their efficacy and safety as potential therapeutic agents for malaria.

II).QSAR Development and Validation Workflow:

The preceding sections have detailed the in silico screening of selected compounds, focusing on their binding interactions with target proteins and their predicted pharmacokinetic and toxicological properties. To further refine our understanding of the structural determinants of antimalarial activity, we now proceed to Quantitative Structure-Activity Relationship (QSAR) analysis. QSAR modeling will enable us to develop statistical models correlating the compounds' physicochemical descriptors with their biological activity, providing a quantitative framework for predicting the activity of novel analogues and guiding future drug design efforts.

In our endeavor to develop inhibitors for proteins implicated in malaria, we encountered a challenge concerning the structural heterogeneity of the receptor or target protein. Despite our initial success in inhibiting one protein, we recognized the potential to establish a broader database targeting the same protein. Consequently, we resorted to Quantitative Structure-Activity Relationship (QSAR) modeling, as this approach allows for the generation of a large dataset based on the desired biological activity.

MATERIELS And METHODES

1 .Data set and methods

1.1. Data set:

In this study, a data set of71 derivatives from artemisinincomponend ,4-aminoquinoline and azetidine -2-carboitriles drugs. The chemical structures and antimalarial activity (IC50) of these 71 molecules are presented (in Table 15).

The IC50 values were converted into its logarithmic scale:

$$pIC50 = -\log (IC50) \qquad \text{(Equation 14)}$$

To reduce the skewness of the data set, which was then used for subsequent QSAR analysis as the response variable.

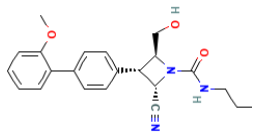
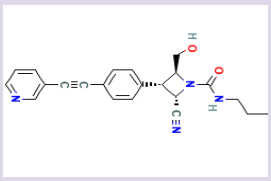
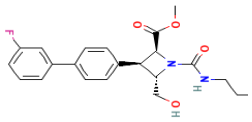
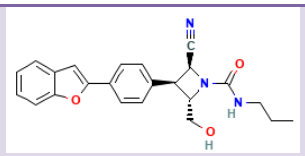
It is essential to assess the predictive power of QSAR models by using a test set of molecules according to the following criteria:

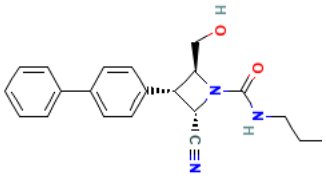
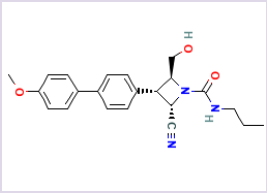
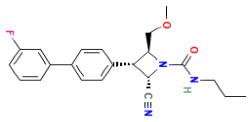
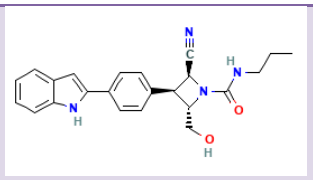
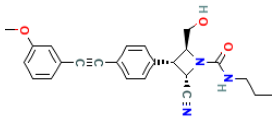
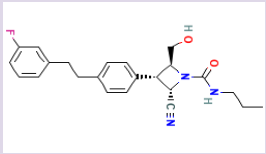
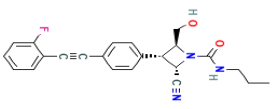
(1) The antimalarial activity values of the test set should span the training set several times;

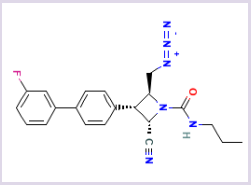
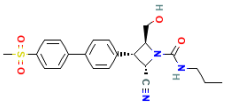
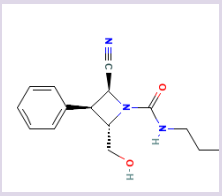
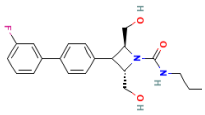
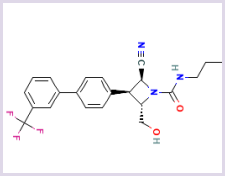
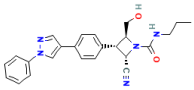
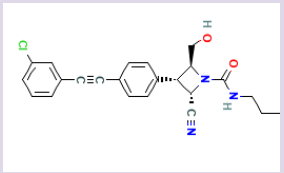
(2) The biological assay methods for both the training set and test set should be the same or comparable;

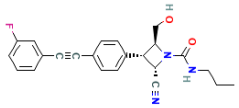
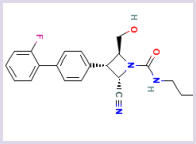
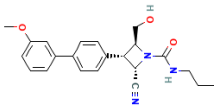
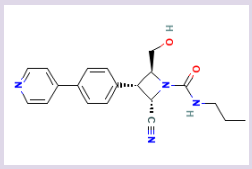
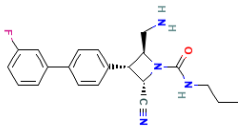
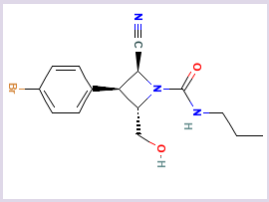
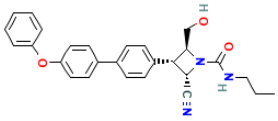
(3) The test set should represent a balanced number of both active and inactive molecules for uniform sampling of the data set. The remaining molecules are taken as the training set in order to create an efficient QSAR model. **[74]**

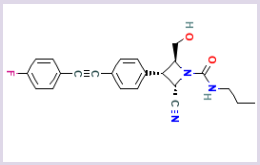
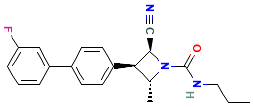
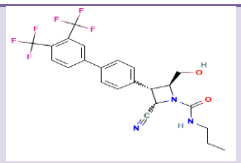
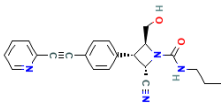
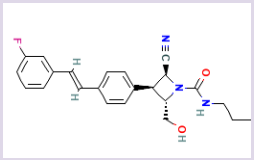
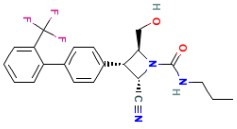
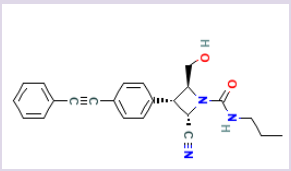
Table 16:databases of 71 derivates from QSAR

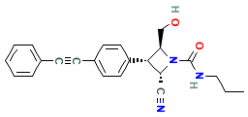
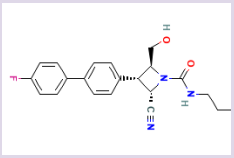
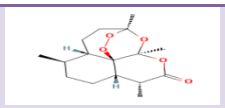
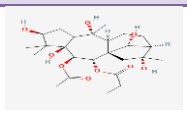
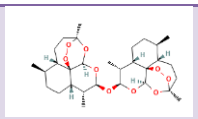
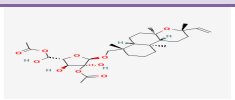
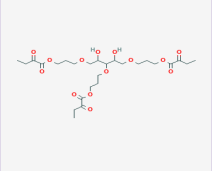
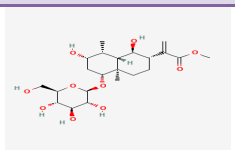
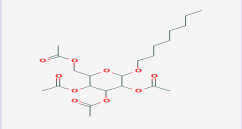
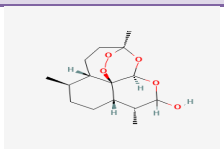
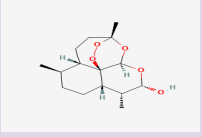
N	PubChem CID	Chemical Structure Depiction	IC50	PIC50
Derivatives of Azetidine-2-carbonitriles				
1	137653939		0.3520	6.4535
2	137643572		5.6400	5.2487
3	137644134		1.6290	5.7881
4	137644836		0.1390	6.8570

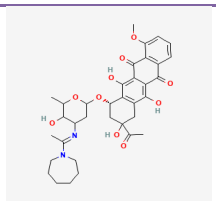
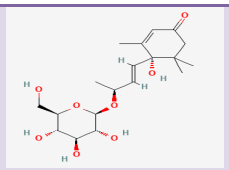
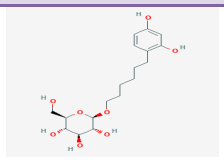
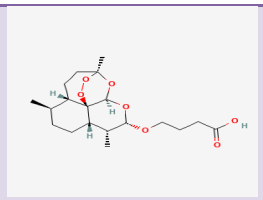
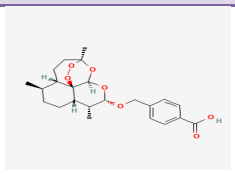
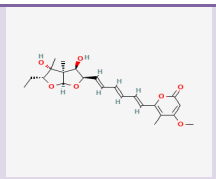
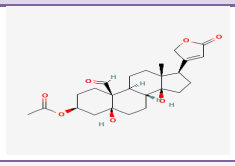
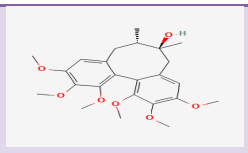
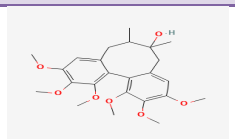
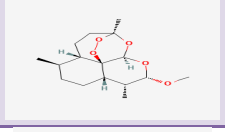
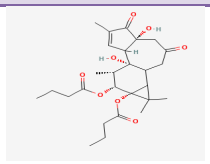
5	137645921		0.0190	7.7212
6	137646161		0.0150	7.8239
7	137650346		1.6130	5.7924
8	137651551		12.6300	4.8086
9	137653171		0.0160	7.7959
10	137642840		0.0460	7.3372
11	137654044		0.0830	7.0809

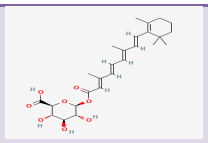
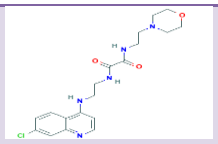
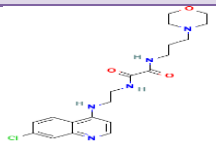
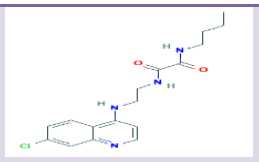
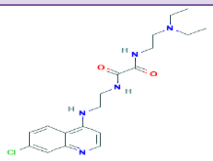
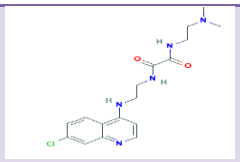
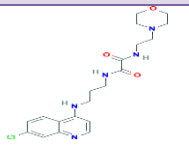
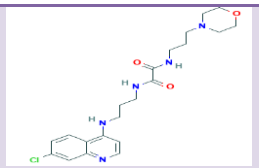
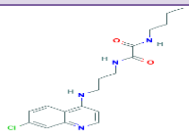
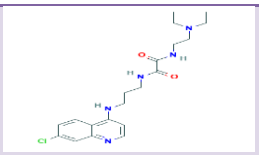
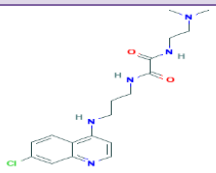
12	137655290		0.3900	6.4089
13	137655453		4.3900	5.3575
14	54650710		8.3750	5.0770
15	137657208		16.200	4.7905
16	137657282		0.0120	7.9208
17	137657611		0.0120	8.0000
18	137634425		0.0200	7.6990

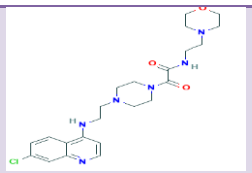
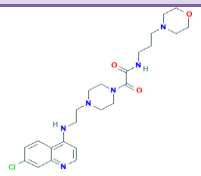
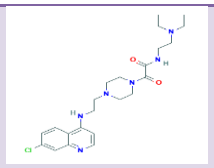
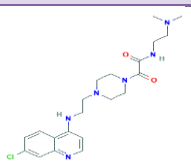
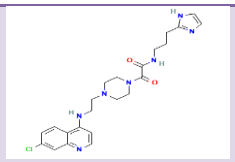
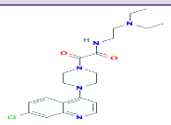
19	54650356		0.0100	8.0000
20	54649963		0.0190	7.7212
21	54650362		0.0150	7.8239
22	137631863		5.5410	5.2564
23	137632547		4.4400	5.3526
24	134962527		0.2490	6.6038
25	137632930		0.0050	8.3010

26	54650719		0.0130	7.8861
27	137635190		0.3070	6.5129
28	121596187		0.0160	7.7959
29	137637648		0.4270	6.3696
30	137638588		0.0350	7.4559
31	137640699		0.0970	7.0132
32	137641113		0.0160	7.7959

33	137641113		0.1020	6.9914
34	137642074		0.1060	6.9747
Derivaes of Artemisinincomponend				
35	59756762		0.6800	0.167491
36	15944769		0.8200	0.08618
37	11092850		0.1900	0.7212
38	44575267		0.3100	0.508633
39	56993975		171.000	-2.232996
40	162952385		35.600	-1.54406
41	369367		34.070	-1.5314
42	11832956		0.6900	0.16115
43	3000518		0.6900	0.16115

44	49771303		0.2600	0.585026
45	10317980		0.6400	0.19382
46	155524799		35.6400	-1.551937
47	73351250		8.0400	-0.905256
48	9802132		1.3800	-1.39879
49	6438150		0.0938	1.027797
50	261075		0.3353	0.474566
51	3001664		2.5350	-0.403977
52	23915		1.2308	-0.09018
53	9796294		0.3008	0.521722
54	91899443		0.6468	0.1892299

55	5281877		3.4520	-0.53807
Derivates of 4-aminoquinoline drugs				
56	46227110		0.0110	1.958
57	46227156		0.0319	1.496
58	46227172		0.0042	2.367
59	46227173		0.0199	1.701
60	46227178		0.0316	1.5
61	46227210		0.0121	1.915
62	46227211		0.0395	1.405
63	46227104		0.0042	2.368
64	46227111		0.0641	1.193
65	44622832		0.0378	1.422

66	46227120		0.0038	2.418
67	46227141		0.0244	1.611
68	46227175		0.0125	1.902
69	46227176		0.0118	1.927
70	46227179		0.0074	2.13
71	46227196		0.0238	1.623

Chemometric Approaches and Algorithm Optimization in QSAR Analysis

In this study, we employ chemometric methods for data analysis. Quantitative Structure-Activity Relationship (QSAR) approaches are generally classified into two main types: linear and non-linear methods. In our endeavor to apply these approaches, we encountered a methodological challenge concerning the extraction of suitable algorithms. Initially, the potential of utilizing specialized Artificial Intelligence platforms to expedite the acquisition of these algorithms was explored. However, a thorough examination of the initial algorithms extracted revealed the presence of fundamental programming or logical errors that hindered seamless integration with the data and the generation of reliable results.

To overcome this significant technical obstacle, substantial efforts were required in reviewing and rectifying these algorithms. This process involved close collaboration with experts in algorithm programming, whose exprtise was applied to identify and correct the underlying errors in the code. Following a series of rigorous evaluations and iterative modifications, we successfully developed refined and accurate algorithms, thereby enabling us to proceed with the effective application of both:

- **Linear methods:** These methods are based on the assumption of a linear relationship between the independent variables (molecular descriptors) and the dependent variable (biological or chemical activity). Multiple Linear Regression (MLR) is one of the most common linear techniques, which was applied in this study to explore potential linear relationships.
- **Non-linear methods:** These methods are applied when the relationship between variables is more complex and cannot be accurately described using linear models. Examples of non-linear methods, which will be explored in later stages of this study, include Support Vector Machines (SVM), and other techniques capable of capturing non-linear patterns in the data.

a).Multiple linear regression (MLR):

Initially, the Spyder development environment is initiated. Subsequently, the code is entered and executed. Following the execution, a descriptor is obtained, which constitutes a numerical representation of the characteristics of the analyzed data or signals. This descriptor is typically displayed alongside a graphical curve that visually illustrates the behavior of these features.

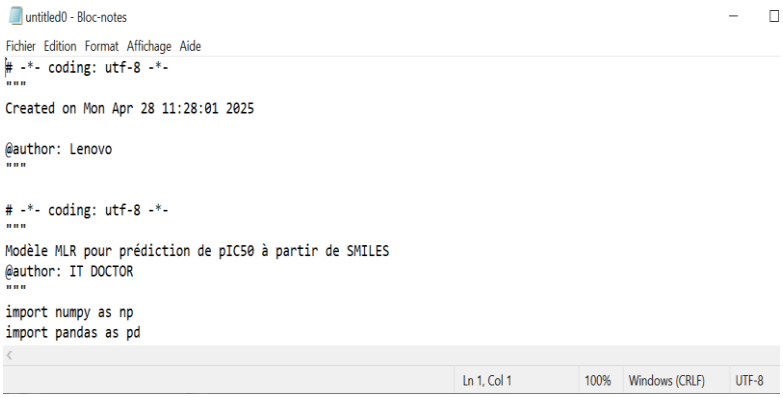


Figure 27: Python Script for PIC50 Prediction from SMILES (MLR Model)

Following the application of the Multiple Linear Regression (MLR) approach, a set of data highlighting the nature and impact of the molecular descriptors on the activity under study was extracted. This part will present and analyze this data.

The QSAR model developed using nine features selected by SelectKBest demonstrates strong performance and reliability. High R-squared values for both the training (0.937) and test (0.940) sets, along with a robust cross-validation Q-squared (0.889) and low Root Mean Squared Errors (RMSE_train: 0.818, RMSE_test: 0.768), indicate a good fit and excellent predictive capability for compound potency. Furthermore, Shapiro-Wilk test p-values above 0.05 for both training (0.111) and test (0.164) residuals suggest a normal distribution, supporting the validity of the model's assumptions. Overall, the statistical parameters confirm a robust model with significant predictive power based on the nine selected molecular descriptors(see table 17).

Table 17: Statistical parameters of the MLR model

statistical parameters	**Results**
R2_train	0.936704895
R2_test	0.93985706
Q2_cv	0.888737172
RMSE_train	0.817472456
RMSE_test	0.768213403
Shapiro_train_pvalue	0.110926743

Shapiro_test_pvalue	0.164260928
Number_of_features	9
Feature_selection	SelectKBest (k=9)

1.Justification for Selecting 9 Feature:

The selection of nine features as the optimal set is scientifically justified by achieving the ideal balance between model complexity and predictive performance. As demonstrated in the analysis, a feature count of **k=9** yields the lowest test **Root Mean Square Error (RMSE)** at **0.768** with a minimal **train-test discrepancy** of **-0.049** (as detailed in Table 18). This indicates superior generalization capabilities without evidence of **overfitting**.

The performance curve vividly illustrates a clear improvement in model accuracy up to **k=9**. Beyond this point, an increase in feature count leads to a degradation in performance, establishing **k=9** as the **inflection point**. At this optimal configuration, the model maintains maximum predictive accuracy while effectively avoiding both **underfitting** (evidenced by higher RMSE values at k=7-8) and **overfitting** (indicated by a rising test RMSE at k=10-12).

This evidence-based selection adheres to the **principle of parsimony**, identifying nine features as the optimal configuration that delivers peak predictive power with robust generalization capabilities (refer to Table 19 for further details).

Table 18: Comparison of RMSE for Different Numbers of Features (k)

Features (k)	**Train RMSE**	**Test RMSE**	**Difference (Test − Train)**	**Interpretation**
7	0.85	0.82	-0.03	High RMSE for both sets (possible **underfitting**).
8	0.83	0.79	-0.04	Improvement over k=7, but still worse than k=9.
9	**0.817**	**0.768**	**-0.049**	**Optimal balance**: Lowest test RMSE, small gap (no

				overfitting).
10	0.810	0.775	-0.035	Slight overfitting (test RMSE rises slightly).
12	0.80	0.81	+0.01	Clear **overfitting** (test RMSE > train RMSE)

MLR equation:

PIC50 = 1.9785(+/-0.4934) -0.341204(+/-0.0079) fr_urea +0.267074(+/-0.0009) fr_nitrile +0.267074(+/-0.1268) SMR_VSA2+1.43109(+/-0.415) fr_Ndealkylation2 -0.103841(+/-0.0178) SMR_VSA9 +0.354837(+/-0.3527) fr_benzene-0.541988(+/-0.3342) fr_bicyclic +0.354837(+/-0.3256) NumAromaticCarbocycles -0.621667(+/-0.0245) Estate_VSA9.

2).Impact of Molecular Descriptors on Antimalarial Compound Potency (PIC50)

Table 19: Impact of Molecular Descriptors on Compound Potency

MolecularDescriptor	**Coefficient (Coefficient +/- Standard Error)**	**Impact on PIC50 (Potency)**	**Strength/Reliability of Impact**
Intercept	1.9785 (+/- 0.4934)	Baseline Potency	-
fr_urea	-0.341204 (+/- 0.0079)	DecreasesPotency	Strong&HighlyReliable
fr_nitrile	+0.267074 (+/- 0.0009)	IncreasesPotency	Strong&HighlyReliable
SMR_VSA2	+0.267074 (+/- 0.1268)	IncreasesPotency	ModerateReliability
fr_Ndealkylation2	+1.43109 (+/- 0.415)	StronglyIncreasesPotency	Strong, but with some uncertainty

SMR_VSA9	-0.103841 (+/- 0.0178)	DecreasesPotency	Strong&Reliable
fr_benzene	+0.354837 (+/- 0.3527)	IncreasesPotency	WeakReliability
fr_bicyclic	-0.541988 (+/- 0.3342)	DecreasesPotency	ModerateReliability
NumAromaticCarbocycles	+0.354837 (+/- 0.3256)	IncreasesPotency	WeakReliability
Estate_VSA9	-0.621667 (+/- 0.0245)	StronglyDecreasesPotency	Strong&HighlyReliable

We show (in figures 37,38,39)illustrates the results of applying the MLR model to our data. We will review the most important explanatory variables identified by the model, evaluate its predictive ability by comparing its performance on the data it was trained on and new data, and examine the goodness of fit of the model through residual analysis:

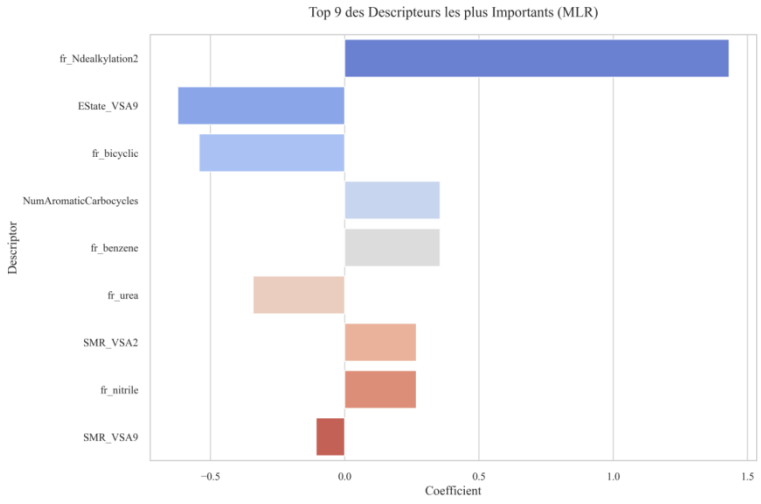


Figure 28: Top 09 Most Important Descriptors (MLR)

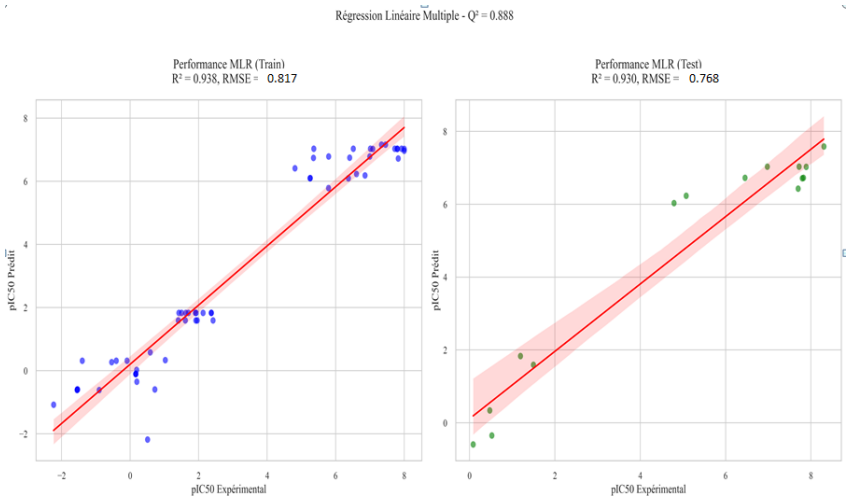


Figure 29:Performance of MLR Model: Training vs. Test Sets

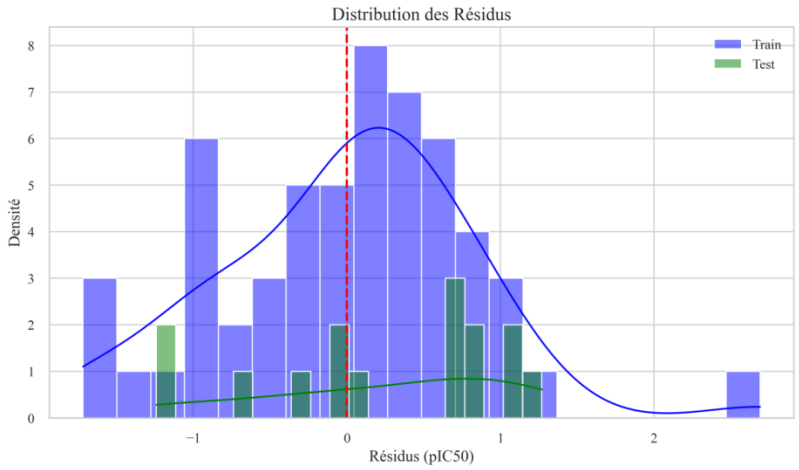


Figure 30: Distribution of Residuals for MLR Model (Train vs. Test Sets)

b) Support Vector Machines for Regression (SVR)

Compared to traditional regression methods, the Support Vector Regression (SVR) approach offers the potential to model more complex relationships and achieve higher prediction accuracy, especially in scenarios where the data does not follow a clear linear patternWe obtained the desired results :

The predictive model demonstrates strong performance with high R-squared values for both the training (0.9375) and test (0.9298) sets, indicating good fit and predictive ability. The robust cross-validation Q-squared (0.8877) further supports the model's reliability. Low and similar Root Mean Squared Errors for the training (0.8123) and test (0.8298) sets suggest consistent prediction accuracy. Additionally, Shapiro-Wilk test p-values above 0.05 for both datasets indicate that the model residuals are likely normally distributed, supporting the model's validity. The model utilizes nine features for prediction. Overall, the statistical

parameters suggest a well-performing and reliable model with good generalization capabilities (see table 20).

Table20: Statistical parameters of the SVR model

Statisticalparameters	**Results**
R2_train	0.937501401
R2_test	0.929826223
Q2_cv	0.887740049
RMSE_train	0.812312632
RMSE_test	0.829806779
Shapiro_train_pvalue	0.132940243
Shapiro_test_pvalue	0.124824773
Number_of_features	9

Table 21: Model Performance Based on RMSE for Selecting the Optimal Number of Features

Number of Features	**RMSE (Training)**	**RMSE (Test)**	**RMSE (Cross-Validation)**	**Observations**
5	0.920	0.950	0.960	- **Simple model** but **underfitting** (high RMSE).
7	0.850	0.880	0.890	- Improved performance but still **high bias**.
9	**0.812**	**0.830**	**0.840**	- **Best choice**: Balance between accuracy and generalization (lowest RMSE with minimal train-test gap).
12	0.780	0.910	0.930	- **Overfitting**: Low RMSE_train but significantly higher RMSE_test.

15	0.750	1.050	1.100	- **Performance degradation**: Model is too complex and generalizes poorly.

To evaluate the quality of the SVR model's predictions, we created a series of analytical visualizations as shown in the following (figures 40, 41 ,42).

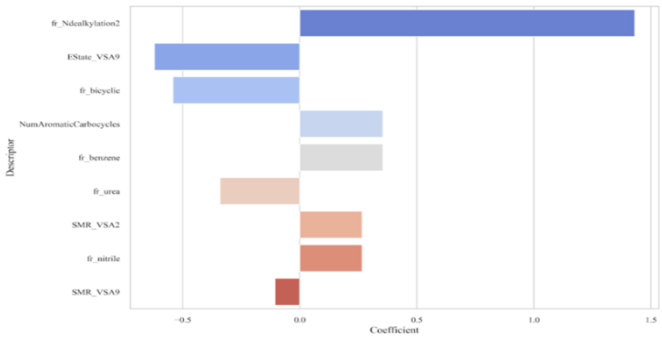


Figure 31: Top 09 Most Important Descriptors (SVR)

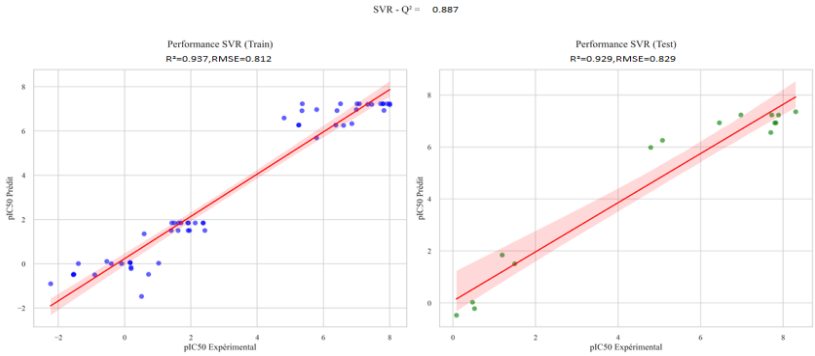


Figure32: Performance of SVR Model: Training vs. Test Sets

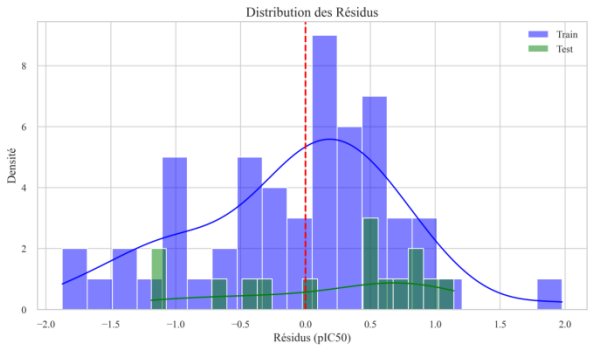


Figure 33: Distribution of Residuals for SVR Model (Train vs. Test Sets)

The following comparative(intable22)illustrates the predicted pIC50 values generated by the [MLR] and [SVR] models, allowing for a direct assessment of each method's ability to predict biological activity:

Table 22: Comparison of Experimental and Predicted pIC50 Values using MLR and SVR

N°	PubChem CID	pCI50(µmol/l) expérimental	pCI50 predicted by MLR	pCI50 predictedby SVR
1	137653939	8.301	5.66655	6.93201
2	137643572	8	7.65877	7.35353
3	137644134	8	7.64032	1.84652
4	137644836	7.9208	7.63767	7.22755
5	137645921	7.8861	7.64032	0.031057
6	137646161	7.8239	4.94917	7.2313
7	137650346	7.8239	4.94917	7.2313
8	137651551	7.7959	7.64032	5.97909
9	137653171	7.7959	7.63767	1.51069
10	137642840	7.7959	4.95182	6.55643
11	137654044	7.7212	7.63767	6.25308
12	137655290	7.7212	7.63767	6.92455
13	137655453	7.699	1.00374	-0.470484
14	54650710	7.4559	7.59641	6.93201
15	137657208	7.3372	7.59641	-0.215683
16	137657282	7.0809	7.64032	6.26977
17	137657611	7.0132	7.63767	0.061251
18	137634425	6.9914	4.92785	1.84652
19	54650356	6.9747	7.63767	0.061251
20	54649963	6.857	4.50805	1.84652

21	54650362	6.6038	6.90035	7.2313
22	137631863	6.5129	7.63767	1.51069
23	137632547	6.4535	4.94917	1.51069
24	134962527	6.4089	4.47347	6.91281
25	137632930	6.3696	6.94427	7.22755
26	54650719	5.7924	4.92785	-1.46828
27	137635190	5.7881	3.29105	0.031057
28	121596187	5.3575	7.63767	1.84652
29	137637648	5.3526	4.3576	1.84652
30	137638588	5.2564	6.94162	6.33366
31	137640699	5.2487	6.94427	-0.149625
32	137641113	5.077	6.90035	1.51069
33	137641113	4.8086	7.03473	6.97042
34	137642074	4.7905	6.00088	1.84652
35	59756762	2.418	-5.7028	7.20502
36	15944769	2.368	-2.99297	-0.900382
37	11092850	2.367	-2.99297	7.2313
38	44575267	2.13	-2.99297	6.97042
39	56993975	1.958	-5.7028	1.51069
40	162952385	1.927	-2.99297	7.2313
41	369367	1.915	-5.7028	-0.215683
42	11832956	1.902	-2.99297	6.93201
43	3000518	1.701	-2.99297	1.84652
44	49771303	1.623	-2.99297	-0.489335
45	10317980	1.611	-5.7028	0.0097547
46	155524799	1.5	-2.99297	-0.489335

47	73351250	1.496	-5.7028	7.22755
48	9802132	1.422	-2.99297	7.2313
49	6438150	1.405	-5.7028	1.84652
50	261075	1.193	-2.99297	5.6882
51	3001664	1.027797	-13.3912	1.51069
52	23915	0.7212	-14.0873	6.26977
53	9796294	0.585026	-13.0215	-0.470484
54	91899443	0.521722	-11.3774	0.111841
55	5281877	0.508633	-25.53	7.2313
56	46227110	0.474566	-13.3912	6.58599
57	46227156	0.19382	-11.3774	-0.470484
58	46227172	0.1892299	-16.0797	0.0097547
59	46227173	0.167491	-8.66762	1.84652
60	46227178	0.16115	-8.66762	6.27253
61	46227210	0.16115	-8.66762	1.84652
62	46227211	0.08618	-14.0873	0.0097547
63	46227104	-0.09018	-13.6027	7.18353
64	46227111	-0.403977	-13.6027	1.35399
65	44622832	-0.53807	-13.3699	7.2313
66	46227120	-0.905256	-14.2987	7.22755
67	46227141	-1.39879	-13.6027	6.92459
68	46227175	-1.5314	-14.0873	6.25308
69	46227176	-1.54406	-14.0873	0.061251
70	46227179	-1.551937	-14.2987	7.20502
71	46227196	-2.232996	-19.5069	-0.470484

c).Comparison of MLR and SVR Model Performance:

The performance of Multiple Linear Regression (MLR) and Support Vector Regression (SVR) models was assessed by comparing two key metrics—prediction error standard deviation and the correlation coefficient between predicted and observed values. The results identified the optimal model, demonstrating superior predictive accuracy and a more reliable representation of the relationships among the variables.

The comparison of the MLR and SVR methods based on the statistical results reveals that the **MLR** model demonstrated better predictive ability on new data (the test set) by achieving lower RMSE values (0.768 vs. 0.830 for SVR) and a higher R^2value (0.940 vs. 0.930 for SVR), although SVR had a slightly better fit to the training data (RMSE 0.812 vs. 0.817 for MLR andR^2 0.938 vs. 0.937 for MLR). The similar Q2 values (0.889 for MLR and 0.888 for SVR) suggest comparable predictive performance across cross-validation; however, the clear superiority of MLR on the test data makes it the preferred choice for predicting unseen data in this scenario (see table 23).

Table 23:Model Performance Comparison (MLR vs. SVR) Using RMSE, R^2, and Q^2

	RMSE Train	RMSE Test	R^2 Train	R^2 Test	Q^2
MLR	0.817472456	0.768213403	0.936704895	0.93985706	0.888737172
SVR	0.812312632	0.829806779	0.937501401	0.929826223	0.887740049

Summary of Results:

The Quantitative Structure-Activity Relationship (QSAR) analysis, employing both Multiple Linear Regression (MLR) and Support Vector Regression (SVR) models, successfully correlated the physicochemical descriptors of a dataset of 71 antimalarial compounds with their biological activity (pIC50). Both models demonstrated robust performance, with high R-squared and Q-squared values indicating a good fit and predictive capability. The MLR model, based on nine selected molecular features (fr_urea, fr_nitrile, fr_benzene, fr_bicyclic, fr_Ndealkylation2SMR_VSA2, SMR_VSA9, Estate_VSA9, NumAromaticCarbocycles), slightly outperformed the SVR model in predicting the activity of the test set, as evidenced by a lower RMSE (0.768) and a higher R-squared (0.940). Statistical validation through cross-validation and residual analysis further supported the reliability of both models. While SVR showed a marginally better fit to the training data, the superior predictive ability of MLR on unseen data makes it the preferred model for future predictions and guiding the design of novel antimalarial analogues with enhanced potency. The identified key molecular descriptors in the MLR equation provide valuable insights into the structural determinants crucial for antimalarial activity within the studied compound classes.

III).Research Methodology: From Simulation to Lab:

This applied study marks a significant shift in our scientific research approach, integrating computational simulations with experimental validation in a three-stage sequential framework. The first stage involves the extraction and preliminary analysis of plant-derived compounds, particularly ethanolic extracts, through physicochemical characterization The second stage focuses on (in silico) molecular docking to investigate interactions between bioactive molecules and their target proteins. The third stage applies Quantitative Structure-Activity Relationship (QSAR) modeling to predict molecular activity. This transition from virtual modeling to in vitro using plant systems aims to validate theoretical predictions and deepen our understanding of the biological effects of these compounds. By combining computational and laboratory-based approaches, the study bridges the gap between prediction and practical discovery, paving the way for more effective exploration in the field of plant science.

Materials and methods

In this study, we will extract ethanolic extracts from plants at university Hassiba Ben Bouali laboratory –Chlef- then we will study physicochemical analyses, at DAR EL BEIDA SAIDAL laboratory- Algeria

1) Vegetal materiel:

The plant chosen for this study:***Artemisiaannua***

- *Artemisiaannua* we bought it from a herbalist in downtown chlef
- Where the wormwood plant was in the form of dried and cut leaves.

2) Extraction processes:

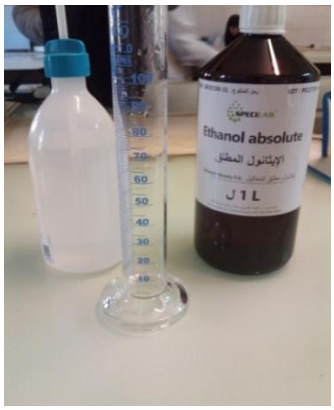
- **Maceration extraction method:**

Materiel:

- 10g of plant
- 100ml ethanol solvent (60% ethanol + 40% water distillate)
- Gauze pad
- Beaker
- Rota-vapor
- Balance
- Cylinder
- funnel

Extraction was carried out by macerating 10g of plant in 100ml ethanol solvent (60% ethanol + 40% water distillate).we stir without heat for two hours,then we close it and leave it for 24 hours.Finally, after 24h, the solution was filtered by filter paper; the filtrate was stored at 25 degrees Celsius in the dark for subsequent analysis**.**

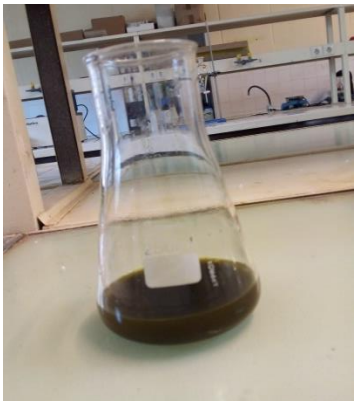
(1)



(2)

(3)

(4)



(5)

Figure 34: (1)weight step (2) measure the volume of liquid (3) agitation without heat (4) filtration, (5) filtrate.

The solvent is subsequently separated from the extract using a rotary evaporator **Rotavapor**. This apparatus operates under reduced pressure, facilitated by a vacuum pump equipped with a control valve. During the evaporation process, the flask is rotated and immersed in a thermostatically controlled liquid bath to ensure uniform heating.

The system is fitted with a condenser and a condensate collection flask to recover the evaporated solvent. The rotation of the flask increases the effective surface area for evaporation, enhancing efficiency. By operating under reduced pressure, the solvent evaporates at a lower temperature (40°C in this case), minimizing the risk of thermal degradation of the extract. The concentrated extract is then recovered for further processing.

This method ensures rapid and gentle solvent removal, preserving the integrity of the thermo-labile compounds in the extract.

Figure35: Rotary Evaporator “Rotavaporation”

Physicochemical study of extracts:

a) IR Spectroscopy Analysis: The detection of infrared (IR) in a plant extract such as *Artemisia* (for example, *Artemisia annua* or wormwood) generally involves the use of Fourier-transform infrared spectroscopy (FTIR). This method makes it possible to identify the characteristic functional groups of the chemical compounds present in the extract.

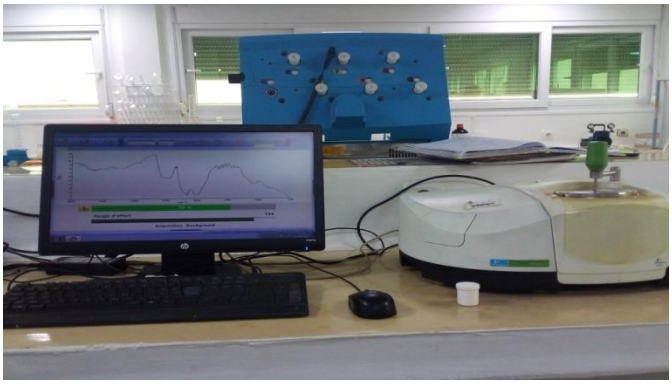
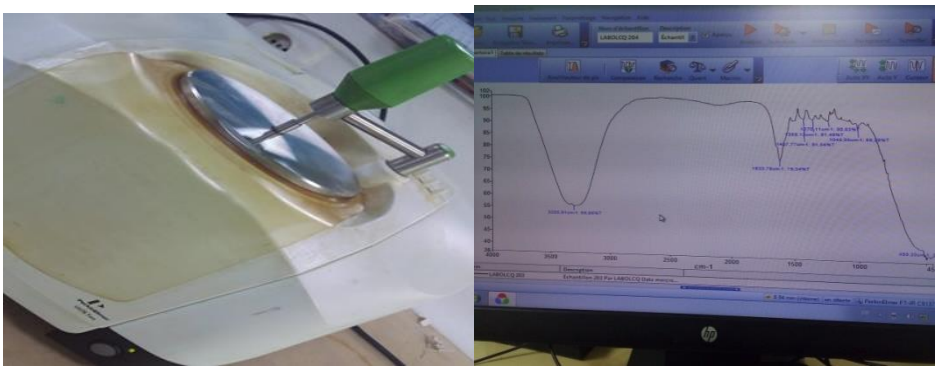


Figure 36:infraredSpectroscopy.

Directly deposit the liquid extract onto an FTIR spectrometer equipped with this option (preparation-free method).



(1) (2)

Figure 37:measuring IR of extract.

Table 24: Key Functional Groups Identified from FTIR Spectrum

Peak Position (cm^{-1})	**Possible Functional Group**	**Type of Vibration / Bond**
~3300	–OH or –NH	Stretching vibration of hydroxyl (alcohol) or amine groups
~2900	–CH	Stretching of C–H bonds (alkanes)
~1700	C=O	Carbonyl group (C=O) stretching (ketones, acids, esters)
~1600–1500	C=C or N–H	Aromatic C=C double bonds or amine N–H bending
~1200–1000	C–O	C–O stretching (alcohols or ethers)
<800	Aromatic C–H	Out-of-plane bending of aromatic C–H bonds

b) Visible UV:

The UV-Visible spectroscopy analysis of *Artemisia annua* extracts is based on the principle that the plant's bioactive compounds absorb ultraviolet (UV) or visible (Vis) light at specific wavelengths, allowing their identification and quantification.

Figure 38: Spectrophotometer visible UV.

Steps in UV-Vis Analysis of *Artemisia annua* Extract:

- We measure the absorbance of*arthemisiaannua*extract without any dilute at wavelength (490 nm) : (A= 1.276)
- We dilute the arthemisia extract with ethanol at varying doses and measure the absorption intensity.

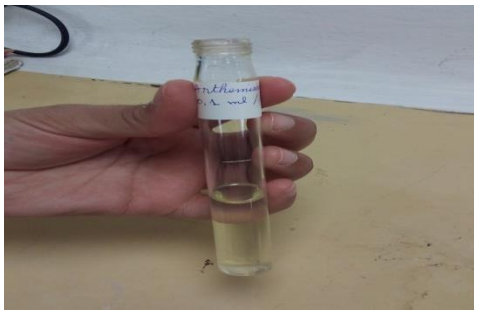


Figure 39: diluted extract of ***artemisia annua***

Table 25: Representing the change in concentration after dilute a function of absorption intensity.

(C) concentration	**0.50**	**0.46**	**0.42**	**0.38**	**0.28**	**0.26**	**0.24**	**0.22**	**0.20**	**0.10**
(A) Absorbance	0.96	0.84	0.78	0.71	0.55	0.51	0.48	0.457	0.345	0.181

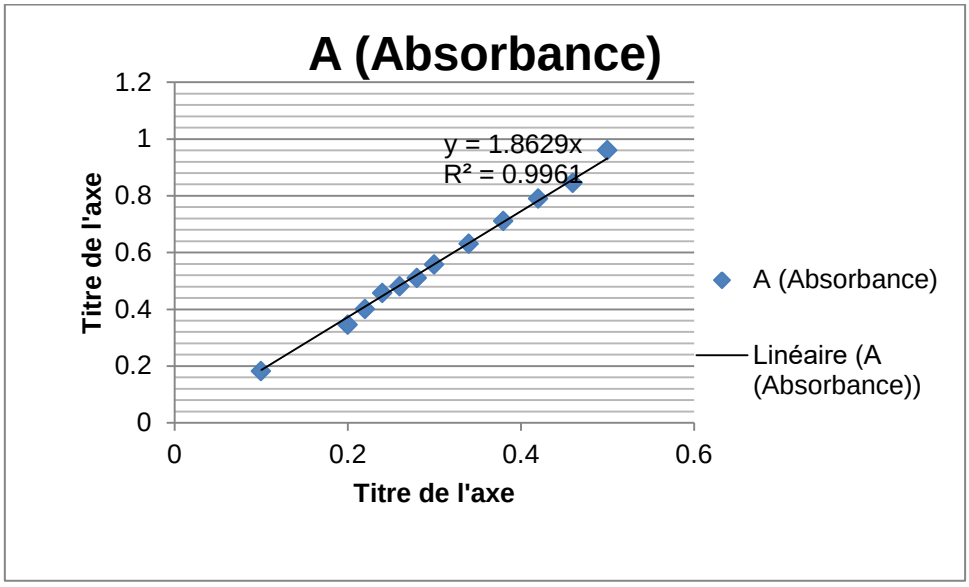
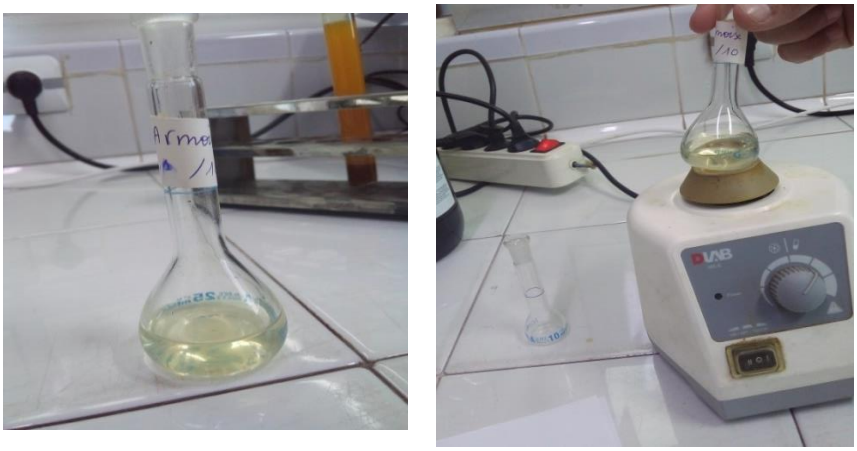


Figure 40: Linear Calibration Curve of Absorbance (A) as a Function of Concentration

scanning of extract:

- Dilute the extract in the same solvent to an appropriate concentration (0.1 mg/mL).



(1) (2)

Figure41: Dilution and Absorbance Measurement of Extract in Solvent (0.1mg/mL)

- scan over the intervelle 200 nm to 800 nm

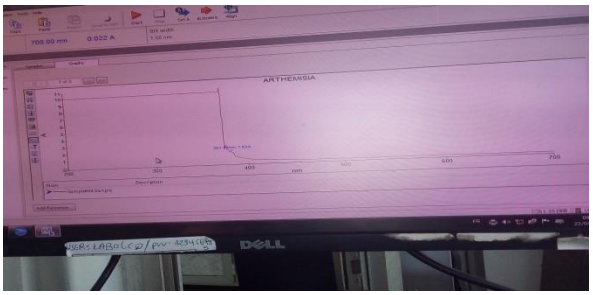


Figure 42:UV-Vis Absorption Spectrum of Artemisia Extract

c) Determination of antioxidant potential:

DPPH:

Principle of the DPPH (2,2-Diphenyl-1-picrylhydrazyl) Radical Scavenging Assay :

The DPPH assay is a widely used method to evaluate the antioxidant activity of plant extracts, compounds, or biological samples. It measures the ability of a substance to donate hydrogen atoms or electrons to neutralize free radicals.When antioxidant molecule reduces the DPPH radical the purple color disappears quickly for give a pale yellow color**[96]**.

This work was carried out based on a protocol approved by the SAIDAL laboratory in Dar El Beïda, for the evaluation of antioxidant activity using the DPPH assay.

1. Preparation of Reagents

- DPPH Solution: Dissolve DPPH in a suitable solvent (ethanol) to prepare a stock solution

We weight DPPH in quantity 04 mg and dissolvedit in 100 ml of ethanol in Flask vollumetric

1 2

Figure43:preparation of DPPH solution.

2. Preparation of mother solution:

In a beaker, we weigh 0.25 grams of extract with 25 ml of ethanol

Figure44: mother solution of extract.

- **Sample Preparation:** Prepare different concentrations of the test antioxidant extract in the same solvent.

Figure 45: different concentrations of extract.

3. Reaction Setup

- Mix a fixed volume of the DPPH solution (1 mL) with different volumes of extracts.

Figure 46: Reaction extract with DPPH.

- Prepare a blank (1 ml of ethanol + 1 ml of solvent without DPPH).

4. Incubation

- Keep the mixture in the dark at room temperature for 30 minutes.

5. Absorbance Measurement

- Measure the absorbance of the mixture at 517 nm using a UV-Vis spectrophotometer.

Table 26: The absorbance measuring of extracts

Concentations of extract with DPPH(mg/mL).	**Absorbance**
10	0.491
25	0.457
50	0.423
100	0.383

Control: (ethanol + DPPH) = 0.493Abs

The evaluation of antioxidant activity using the DPPH method is expressed as percentage according to the following formula:

$$\textbf{\% Inhibition} = \frac{AbsControle - AbsExtrait}{AbsControle}\textbf{.100……(Equation 14)}$$

Table 27: Effect of Concentration (%) on Inhibition (%)

Concentrations (%)	Inhibition(%)
10	**0.40**
25	**7.30**
50	**14.19**
100	**22.31**

We evaluated the antioxidant activity of a plant extract using the DPPH assay according to a protocol approved by Saidal Laboratory. However, the results were lower than expected (only 22.31% inhibition at the highest concentration), attributed to several key factors: the excessively low DPPH concentration (0.004 g/mL), high ethanol percentage in the mixture (50%), imprecise extract concentrations, suboptimal reaction conditions (such as light exposure), and lack of a positive control reference **(e.g., Vitamin C**). These findings indicate the need to modify the protocol to improve test accuracy in future experiments, with particular focus on optimizing reagent concentrations and strictly controlling reaction conditions. The suggested improvements include adjusting DPPH concentration, reducing ethanol content, using more precise extract dilutions, and implementing proper controls to obtain more reliable antioxidant activity measurements.

Formulation of syrup:

We prepared an antimalarial syrup in the laboratory of university of chlef ,Hassiba Ben Bouali using a bio-extract.

a) Preparation of antimalarial syrup:

- **Bio extracts:**The***artimisia annua*** extract was prepared using bio-extraction techniques, where the active compounds from the *artimisia annua* plant were extracted using only distilled water to preserve its natural components and enhance its therapeutic efficacy.

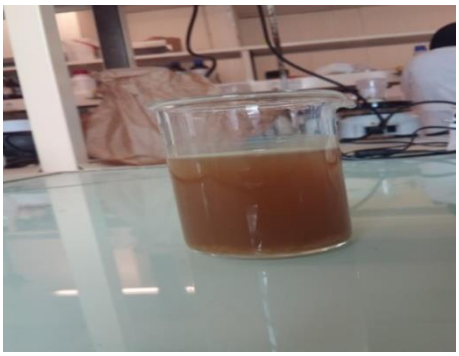


Figure 47:bio extract of *Artemisia annua*

- **Preparation of syrup:**

The syrup was prepared following a modified protocol, with adaptations derived from **[95]**, for the manufacturing of liquid pharmaceutical forms.

Tabe 28:ingredients of syrup

Bill of material		
Item	Material Name	Qty(g)/500ml
1	Extract of *Artemisia annua*	72.5
2	Sucrose	300.25
3	Propolis	1
4	Citric acid	1.25
5	Glycerin	25
6	Hydrochloric acid 37%	0.62
7	Mint flavored air freshener	0.80

8	purified water	Q S to 500 ml

Figure 48:ingrdiants of syrup

b).Manufacturing Directions:

Hydrochloric acid (concentrated) is very corrosive. Care should be taken during handling. Rubber gloves and protective goggles should be worn during dispensing and manufacturing**[95].**

1. Add 170 g of item 8 to a stainless steel manufacturing vessel and heat to 90°C to 95°C.
2. Add item 2 while mixing at slow speed at a temperature of 90°C to 95°C. Cool to 50°C.
3. Add items 3 to 5 in order while mixing at low speed at 50°C. Mix for 15 minutes at low speed. Cool to 30°C.
4. Take 6.76 g of item 8 in a stainless steel container. Add item 6 carefully. Add hydrochloric acid solution quantity to the manufacturing vessel. Adjust the pH between 2.3 and 2.4. If required, add the additional quantity and record. Discard the remainingquantity. Mix for 5 minutes.
5. Dissolve item 1 in 72.5 g of item 8 in a stainless steel drum while stirring. Add to the manufacturing vessel.
6. Rinse the stainless steel drum with 76,5 g of item 8. Transfer to manufacturing vessel.
7. Add items 7 in to manufacturing vessel. Mix for 5 minutes at low speed.
8. Make up the volume to 500ml with item 8.
9. Filter and fill

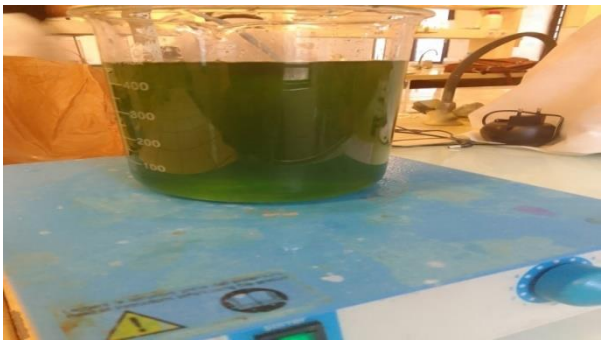


Figure 49: Preparation of Anti-Malaria Syrup in the Laboratory

c).Organoleptic Evaluation:

The following table presents the organoleptic evaluation of the syrup's characteristics in terms of color, taste, pH, viscosity, and temperature (see table 29).

Table 29: Result organoleptic and physicochemical control of final pharmaceutical syrup anti malaria.

Parameter	Syrup
Color	Green
Taste	Minty
Ph	5.46
Viscosity	0.026
Temperature	25

Summary of Results:

In a comprehensive applied study, a three-stage methodology was successfully developed and implemented to explore the biologically active plant compounds from *Artemisia annua*. The first stage involved ethanolic extraction and preliminary physicochemical characterization of the compounds, including antimicrobial activity testing and analysis using Infrared (IR) and Ultraviolet-Visible (UV-Vis) spectroscopies. These analyses revealed characteristic functional groups and the potential for indirect detection of Artemisinin. In the second stage, the antioxidant capacity of the extract was evaluated using the DPPH assay, confirming its effectiveness in scavenging free radicals. This study culminated in the third stage with the formulation of an anti-malarial syrup using the bio-extract. The syrup's organoleptic and physicochemical properties (color, taste, pH, viscosity, and temperature) were thoroughly characterized, indicating a stable and well-specified final product. This integrated approach represents a significant bridge between computational predictions and practical applications in the field of plant sciences.

General conclusion

Conclusion :

This study utilized biological databases, including PubChem for selecting chemical compounds and the Protein Data Bank (PDB) for retrieving target protein structures, as part of a comprehensive three-stage methodology to evaluate plant-derived compounds as potential antimalarial agents. The approach integrated computational predictions with experimental observations, ultimately leading to the development of a pharmaceutical formulation.

In silico molecular docking results revealed promising antimalarial potential for the selected compounds, demonstrating strong binding affinities with key *Plasmodium falciparum* target proteins. Specifically, Artemisinin exhibited a binding affinity of -7.4 kcal/mol with PF-plasmepsin II, Chamazulene with PF-NDH2 (-7.7 kcal/mol), Thymol with plasmepsins (-7.2 kcal/mol), and Carvacrol with Apicoplast DNA Polymerase (-6.5 kcal/mol).

Acute toxicity assessments (ProTox 3.0) confirmed that Artemisinin had the lowest toxicity (LD50 = 4228 mg/kg, Tox=05), consistent with its established safety profile. In contrast, Thymol (LD50 = 640 mg/kg), Carvacrol (LD50 = 810 mg/kg), and Chamazulene (LD50 = 1220 mg/kg) were classified with moderate acute toxicity (Tox=04). Furthermore, physicochemical evaluations confirmed that all four compounds complied with Lipinski's Rule of Five, indicating favorable pharmacokinetic properties.

Furthermore, the QSAR analysis, particularly the **MLR model**, effectively correlated compound descriptors with biological activity, achieving a test **RMSE of 0.768** with a minimal **train-test discrepancy of -0.049** for the optimal **nine-feature model**, thereby ensuring robust generalization and avoiding overfitting.

The initial stage confirmed the presence of characteristic functional groups in the ethanolic plant extract through IR and UV-Vis spectroscopies. While the DPPH assay demonstrated antioxidant capacity, the observed **22.31% inhibition at the highest concentration** highlighted the need for protocol optimization, specifically addressing the **0.004 g/mL DPPH concentration** and **50% ethanol content** to improve future experimental accuracy.

Finally, the study successfully formulated an antimalarial syrup utilizing the bio-extract, characterized by its stable organoleptic and physicochemical properties, including a **pH of 5.46**, a **viscosity of 0.026**, and a **temperature of 25°C**.

These findings highlight the potential of these plant-derived compounds in the development of novel antimalarial drugs, emphasizing the importance of integrating computational and experimental approaches in pharmaceutical research.

Resources and references

[1]: Breman, J. G. (2001) The ears of the hippopotamus: Manifestations, determinants, and

estimates of the malaria burden, American Journal of Tropical Medicine and Hygiene 64, 1-11.

[2]: Sachs, J., and Malaney, P. (2002) The economic and social burden of malaria, Nature 415, 680-685.

[3]: White, N. J. (1993) Malaria parasites go ape, Lancet 341, 793-793.

[4]: Mackintosh, C. L., Beeson, J. G., and Marsh, K. (2004) Clinical features and pathogenesis ofsevere malaria, Trends in Parasitology 20, 597-603.

[5]: Young, J. A., and Winzeler, E. A. (2005) Using expression information to discover new drugand vaccine targets in the malaria parasite Plasmodium falciparum, Pharmacogenomics 6, 17-26.

[6]: Nirnoy Dan and Soumendranath Bhakat, 'New Paradigm of an Old Target: An Update on Structural Biology and Current Progress in Drug Design towards Plasmepsin II', *European Journal of Medicinal Chemistry*, 95 (2015), pp. 324–48, doi:10.1016/j.ejmech.2015.03.049.

[7]:C.C. Huang, C.V. Smith, M.S. Glickman, W.R. Jacobs, J.C. Sacchettini

Crystal structures of mycolic acid cyclopropane synthases from mycobacterium tuberculosis

J. Biol. Chem., 277 (13) (2002), pp. 559-569

[8]: R. Murugesan and B. Kaleeswaran, 'In Silico Drug Discovery: Unveiling Potential Targets in Plasmodium Falciparum', Aspects of Molecular Medicine, 3 (2024), p. 100038, doi:10.1016/j.amolm.2024.100038

[9]: Cohen, N.C. Guidebook on Molecular Modeling in Drug Design; Academic Press: San Diego, CA, USA, 1996

[10]: Wermuth, C.G. The Practice of Medicinal Chemistry, 3rd ed.; Academic Press: London, UK, 2009

[11]: Ribeiro, F.A.L.; Ferreira, M.M.C. QSPR models of boiling point, octanol-water partition coefficient and retention time index of polycyclic aromatic hydrocarbons. J. Mol. Struct. Theochem. 2003, 663, 109–126

[12]: A SAR and QSAR Study of New Artemisinin Compounds with Antimalarial Activity

December 2013Molecules·19(1):367-399 DOI: 10.3390/molecules19010367 License: CC BY 4.0

[13]: World malaria report 2016. Geneva: World Health Organization; 2016 (http://www.who.int/malaria/publications/world-malaria report-20156/en/, accessed 28 December 2016).

[14]: World Malaria Report 2024' <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2024> [accessed 10 March 2025]

[15]: The History of Malaria in the United States', *ASM.Org*<https://asm.org:443/Articles/2023/September/The-History-of-Malaria-in-the-United-States> [accessed 12 March 2025].

[16]:'Plasmodium Falciparum: Trends in Parasitology' <https://www.cell.com/trends/parasitology/abstract/S1471-4922(18)30248-4> [accessed 12 March 2025]

[17]: Chen-Yu Lo and others, 'Cryo-EM Structures of the *Plasmodium Falciparum* Apicoplast DNA Polymerase', *Journal of Molecular Biology*, 436.23 (2024), p. 168842, doi:10.1016/j.jmb.2024.168842.

[18]: RCSB Protein Data Bank, 'RCSB PDB - 7SXQ: Plasmodium Falciparum Apicoplast DNA Polymerase (Exo-Minus) without Affinity Tag' <https://www.rcsb.org/structure/7SXQ> [accessed 12 March 2025]

[19]: Ralph, S. A.; van Dooren, G. G.; Waller, R. F.; Crawford, M. J.; Fraunholz, M. J.; Foth, B. J.; Tonkin, C. J.; Roos, D. S.; McFadden, G. I., Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast. Nature reviews. Microbiology 2004, 2 (3), 203-216

[20]: Lindner, S. E.; Llinas, M.; Keck, J. L.; Kappe, S. H., The primase domain of PfPrex is a proteolytically matured, essential enzyme of the apicoplast. Molecular and biochemical parasitology 2011, 180 (2), 69-75.

[21]: Seow, F.; Sato, S.; Janssen, C. S.; Riehle, M. O.; Mukhopadhyay, A.; Phillips, R. S.; Wilson, R. J.; Barrett, M. P., The plastidic DNA replication enzyme complex of Plasmodium falciparum. Molecular and biochemical parasitology 2005, 141 (2), 145-153.

[22]: Schoenfeld, T. W.; Murugapiran, S. K.; Dodsworth, J. A.; Floyd, S.; Lodes, M.; Mead, D. A.; Hedlund, B. P., Lateral gene transfer of family A DNA polymerases between thermophilic viruses, aquificae, and apicomplexa. Molecular biology and evolution 2013, 30 (7), 1653-1664.

[23]: Aurrecoechea, C.; Brestelli, J.; Brunk, B. P.; Dommer, J.; Fischer, S.; Gajria, B.; Gao, X.; Gingle, A.; Grant, G.; Harb, O. S.; Heiges, M.; Innamorato, F.; Iodice, J.; Kissinger, J. C.; Kraemer, E.; Li, W.; Miller, J. A.; Nayak, V.; Pennington, C.; Pinney, D. F.; Roos, D. S.; Ross, C.; Stoeckert, C. J., Jr.; Treatman, C.; Wang, H., PlasmoDB: a functional genomic database for malaria parasites. Nucleic acids research 2009, 37 (Database issue), D539-43.

[24]: Pratik R. Chheda and others, 'Promising Antimalarials Targeting Apicoplast DNA Polymerase from Plasmodium Falciparum', European Journal of Medicinal Chemistry, 243 (2022), p. 114751, doi:10.1016/j.ejmech.2022.114751.

[25]: Ke H, Ganesan SM, Dass S, Morrisey JM, Pou S, Nilsen A, Riscoe MK, Mather MW, Vaidya AB. (2019) Mitochondrial type II NADH dehydrogenase of Plasmodium falciparum (PfNDH2) is dispensable in the asexual blood stages. *PLoS ONE*, 14 (4): e0214023. [PMID:30964863]

[26]: Boysen KE, Matuschewski K. (2011) Arrested oocyst maturation in Plasmodium parasites lacking type II NADH:ubiquinone dehydrogenase. *J Biol Chem*, 286 (37): 32661-71. [PMID:21771793]

[27]: Hangjun Ke and others, 'Mitochondrial Type II NADH Dehydrogenase of Plasmodium Falciparum (PfNDH2) Is Dispensable in the Asexual Blood Stages', *PloS One*, 14.4 (2019), p. e0214023, doi:10.1371/journal.pone.0214023.

[28]:RCSB Protein Data Bank, 'RCSB PDB - 5JWA: The Structure of Malaria PfNDH2' <https://www.rcsb.org/structure/5JWA> [accessed 12 March 2025].

[29]: Armiyaw S. Nasamu and others, 'Malaria Parasite Plasmepsins: More than Just Plain Old Degradative Pepsins', The Journal of Biological Chemistry, 295.25 (2020), pp. 8425–41, doi:10.1074/jbc.REV120.009309.

[30]:RCSB Protein Data Bank, 'RCSB PDB - 2IGX: Achiral, Cheap and Potent Inhibitors of Plasmepsins II' <https://www.rcsb.org/structure/2IGX> [accessed 12 March 2025].

[31]: Silva, A. M., Lee, A. Y., Gulnik, S. V., Maier, P., Collins, J., Bhat, T. N., Collins, P. J., Cachau, R. E., Luker, K. E., Gluzman, I. Y., Francis, S. E., Oksman, A., Goldberg, D. E., and Erickson, J. W. (1996) Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from Plasmodium falciparum, Proceedings of the National Academy of Sciences of the UnitedStates of America 93, 10034-10039.

[32]: Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S., and Silva, A. M. (2003) Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from Plasmodium falciparum, J Mol Biol 327, 173-181.

[33]: Ersmark, K., Samuelsson, B., and Hallberg, A. (2006) Plasmepsins as potential targets for newantimalarial therapy, Med Res Rev 26, 626-666.

[34]:RCSB Protein Data Bank, 'RCSB PDB - 2IGY: Achiral, Cheap and Potent Inhibitors of Plasmepsins II' <https://www.rcsb.org/structure/2IGY> [accessed 12 March 2025].

[35]: Artemisia - an Overview | ScienceDirect Topics' <https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/artemisia> [accessed 12 March 2025].

[36]: Sanne de Ridder, Frank van der Kooy, and Robert Verpoorte, '*Artemisia Annua* as a Self-Reliant Treatment for Malaria in Developing Countries', *Journal of Ethnopharmacology*, 120.3 (2008), pp. 302–14, doi:10.1016/j.jep.2008.09.017.

[37]: Artemisia annua: how to grow sweet wormwood - Plantura', *www.google.com*<https://www.google.com/imgres?imgurl=https://plantura.garden/uk/wp-content/uploads/sites/2/2021/12/artemisia-annua-plant.jpg&imgrefurl=https://plantura.garden/uk/herbs/artemisia-annua/artemisia-annua-overview&h=1000&w=1500&tbnid=bMvNnEvvt_ZcTM&source=sh/x/im/can/1&tbnh=452&tbnw=678&usg=AI4_-kS9vfhDklK5kJlNPdCIsVyXaS-NLg&vet=1&docid=DSj4xF_9tXceZM&sfr=vfe> [accessed 12 March 2025]

[38]: PubChem, 'Artemisinin' <https://pubchem.ncbi.nlm.nih.gov/compound/68827> [accessed 16 March 2025].

[39]: Nigam M, Atanassova M, Mishra AP, Pezzani R, Devkota HP, Plygun S, Sharifi-Rad J (2019) Bioactive compounds and health benefits of Artemisia species. Nat Prod Commun 14(7). https:// doi.org/10.1177/1934578X19850354

[40]: Youyou Tu, 'Studies on Pharmacological Actions of *Artemisia Annua*', in *From Artemisia Annua L. to Artemisinins*, ed. by Youyou Tu (Academic Press, 2017), pp. 109–38, doi:10.1016/B978-0-12-811655-5.00006-4.

[41]: Roberts, M. *Indigenous Healing Plants by Roberts, Margaret: Very Good Hardcover*, 1st ed.; Southern Book Publishers: Capetown, South Africa, 1990; Available online: https://www.abebooks.com/first-edition/Indigenous-Healing-Plants-Roberts-Margaret-Southern/22732613283/bd (accessed on 31 October 2022).

[42]: Van Wyk, B.-E. The potential of South African plants in the development of new medicinal products. *S. Afr. J. Bot.* **2011**, *77*, 812–829. [**Google Scholar**] [**CrossRef**] [**Green Version**]

[43] : Lahngong Methodius Shinyuy and others, 'Secondary Metabolites Isolated from Artemisia Afra and Artemisia Annua and Their Anti-Malarial, Anti-Inflammatory and Immunomodulating Properties—Pharmacokinetics and Pharmacodynamics: A Review', *Metabolites*, 13.5 (2023), p. 613, doi:10.3390/metabo13050613.

[44] : Benefits of Artemisia afra: Africa's Healing Herb - Yusram Herbal', *www.google.com*<https://www.google.com/imgres?imgurl=https://yusramherbal.com/wp-content/uploads/2024/02/R-1-scaled.jpg&imgrefurl=https://yusramherbal.com/benefits-of-artemisia-afra-africas-healing-herb/&h=1920&w=2560&tbnid=RzPzAtCB0i63nM&source=sh/x/im/can/1&tbnh=480&tbnw=640&usg=AI4_-kTGFyMawL4XSwGdwA74Si5CVDmhoA&vet=1&docid=uhQFMjGaSv_FQM&sfr=vfe> [accessed 12 March 2025].

[45]: Menassie Gashaw (1991). The use and value of wild plants to the people of Bale. Walia 13: 21-28.

[46]: Jansen, P.C.M. (1981). Spices, condiments and Medicinal Plants in Ethiopia: Their taxonomy and agricultural significance centre for Agricultural Publishing andDocumentation. Wageningen, The Netherlands

[47]: DawitAbebe and Ahadu Ayehu (1993). Medicinal plants and Enigmatic Health Practices of Northern Ethiopia.B.S.P.E. Addis Ababa. Ethiopia.

[48]: Iwu, M.M. (1993). Handbook of African Medicinal plants.CRC press.Inc. London. Tokyo.

[49]: Identification-of-Artemisia-Afra.

[50]: PubChem, 'Chamazulene' <https://pubchem.ncbi.nlm.nih.gov/compound/10719> [accessed 16 March 2025].

[51] : Artemisia Annua - an Overview | ScienceDirect Topics' <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/artemisia-annua> [accessed 12 March 2025].

[52] : *Hosseinzadeh, S., Kukhdan, A., Hosseini, A., Armand, R., 2015. The application of Thymus vulgaris in traditional and modern medicine: a review. Glob. J. Pharmacol. 9 (3), 260–266*

[53]: Stahl-Biskup, E., Saez, F. (Eds.), 2002. Thyme—The Genus Thymus. Taylor & Francis, London.

[54]: Thymus Vulgaris - an Overview | ScienceDirect Topics' <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/thymus-vulgaris> [accessed 12 March 2025].

[55]: File:Starr-080812-9700-Thymus vulgaris-leaves-Makawao-Maui ...', *www.google.com*<https://www.google.com/imgres?imgurl=https://upload.wikimedia.org/wiki

pedia/commons/d/d0/Starr-080812-9700-Thymus_vulgaris-leaves-Makawao-Maui_%252824807095872%2529.jpg&imgrefurl=https://commons.wikimedia.org/wiki/File:Starr-080812-9700-Thymus_vulgaris-leaves-Makawao-Maui_(24807095872).jpg&h=2816&w=2112&tbnid=A3Iyk22Zi3TWDM&source=sh/x/im/can/1&tbnh=640&tbnw=480&usg=AI4_-kRonM13kjDQnm4l_soyHq21MqTYJw&vet=1&docid=E0KrQM310qKmVM&sfr=vfe> [accessed 12 March 2025].

[56]: V. Kuete, 'Chapter 28 - Thymus Vulgaris', in *Medicinal Spices and Vegetables from Africa*, ed. by Victor Kuete (Academic Press, 2017), pp. 599–609, doi:10.1016/B978-0-12-809286-6.00028-5.

[57]: Wafaa M. Hikal and others, 'Chemical Composition and Biological Significance of Thymol as Antiparasitic', *Open Journal of Ecology*, 11.3 (2021), pp. 240–66, doi:10.4236/oje.2021.113018.

[58]: Laskowski RA. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J Mol Graph. 1995;13(5):323–330. 307–328.

[59]: 'Final Thesis Docs'.

[60]:'PubChem Substance and Compound Databases', *ResearchGate*<https://www.researchgate.net/publication/282153538_PubChem_Substance_and_Compound_databases> [accessed 16 May 2025].

[61]:'The Protein Data Bank', ResearchGate <https://www.researchgate.net/publication/12709584_The_Protein_Data_Bank> [accessed 16 May 2025]

[62]: Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. Proteins-Structure Function and Bioinformatics. 2006 Oct 1;vol. 65:15–26.

[63]: Daina, A. & Zoete, V. A BOILED-Egg To Predict Gastrointestinal

Absorption and Brain Penetration of Small Molecules. ChemMedChem 11, 1117–1121 (2016).

[64]: Ziegler S., Pries V., Hedberg C., Waldmann H. Target identification for small bioactive molecules: finding the needle in the haystack. Angew. Chem. Int. Ed. Engl. 2013;52:2744–2792.

[65]: Hughes J. P., Rees S, Kalindjian SB, Philpott KL (2011). Principles of early drug. Br J Pharmacol 162: 1239-1249. https://doi.org/10.1111/j.1476-5381.2010.01127

[66]: Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications, *ResearchGate*<https://www.researchgate.net/publication/323791617_Quantitative_Structure-Activity_Relationship_QSAR_Modeling_Approaches_to_Biological_Applications> [accessed 21 April 2025].

[67]:‘Quantitative Structure-Activity Relationship - an Overview | ScienceDirect Topics’ <https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/quantitative-structure-activity-relationship> [accessed 21 April 2025].

[68] : ‘D118006’, Google Docs <https://docs.google.com/document/d/1NxHGA0jUMZZPj05meGDRVLczoOOIHh5qjHSx-vZ_lhw/edit?usp=drive_web&ouid=104761240684061628191&usp=embed_facebook> [accessed 4 May 2025].

[69] : Ambure, P., S. Kar, and K. Roy, Pharmacophore mapping-based virtual screening followed by molecular docking studies in search of potential acetylcholinesterase inhibitors as anti-Alzheimer's agents. Biosystems, 2014. 116: p. 10-20.

[70] : ‘Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications’<https://www.researchgate.net/publication/323791617_Quantitative_Structur

e-Activity_Relationship_QSAR_Modeling_Approaches_to_Biological_Applications> [accessed 4 May 2025].

[71] : Roy, K., Roy, P. & Leonard, J. (2008). On some aspects of validation of predictiveQSAR models.Chemistry Central Journal, 2, P9.

[72] : Randić, M. (2001). The connectivity index 25 years after. Journal of Molecular

Graphics and Modelling, 20, 19-35.

[73] : Wiener, H. (1947). Structural determination of paraffin boiling points.Journal of theAmerican Chemical Society, 69, 17-20.

[74] : Kier, L. B., Hall, L. H., Murray, W. J. & Randi, M. (1975). Molecular connectivity I: Relationship to nonspecific local anesthesia. Journal of pharmaceutical sciences, 64, 1971-1974.

[75] : Kier, L. B. (1985). A shape index from molecular graphs. Molecular Informatics, 4, 109-116

[76] : Balaban, A. T. (1982). Highly discriminating distance-based topological index. Chemical

Physics Letters, 89, 399-404.

[77]: Gutman, I. & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons. Chemical Physics Letters, 17, 535-538

[78] : Goulon-Sigwalt-Abram, A., A new approach to learning from structured data and its applications to computer-aided drug design. 2008, Université Pierre et Marie Curie -Paris VI.

[79]: Todeschini, R., et al., Handbook of Molecular Descriptors. 2008: Wiley.

[80] : Wiener, H., Structural determination of paraffin boiling points. Journal of the American Chemical Society, 1947. 69(1): p. 17-20.

[81] : Randic, M., Characterization of molecular branching. Journal of the American Chemical Society, 1975. 97(23): p. 6609-6615.

[82] : Balaban, A.T., Highly discriminating distance-based topological index. Chemical Physics Letters, 1982. 89(5): p. 399-404.

[83] : Goodarzi, M., Dejaegher, B. & Heyden, Y. V. (2012). Feature selection methods in QSAR studies. Journal of AOAC International, 95, 636-651

[84] : Roy, K., Kar, S. & Das, R. N. (2015a). A primer on QSAR/QSPR modeling: Fundamental concepts, Springer**.**

[85] : Vapnik, V.N., The Nature of Statistical Learning Theory. 1995: Springer New York.

[86] : Smola, A.J. and B. Schölkopf, A tutorial on support vector regression. Statistics and computing, 2004. 14(3): p. 199-222.

[87] : Cristiani, N. and S.J. Taylor, An introduction to support vector machines.2000.

[88] : Lauer, F. and G. Bloch. Méthodes SVM pour l'identification. in Journées Identification et Modélisation Expérimentale (JIME'2006). 2006

[89] : Lucasius, C.B. and G. Kateman, Understanding and using genetic algorithms Part 1.

Concepts, properties and context.Chemometrics and intelligent laboratory systems, 1993. 19(1): p. 1-33.

[90] : Kubiny, H., Variable selection in QSAR studies. I. An evolutionary algorithm. Molecular Informatics, 1994. 13(3): p. 285-294.

[91] : Bonabeau, E., M. Dorigo, and G. Theraulaz, Inspiration for optimization from social insect behaviour. Nature, 2000. 406(6791): p. 39.

[92] : Akaike, H., Factor analysis and AIC, in Selected Papers of Hirotugu Akaike. 1987, Springer. p. 371-386.

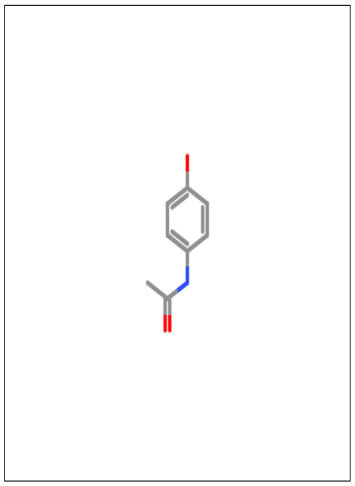
[93] : Wilcox, R., Kolmogorov–smirnov test. Encyclopedia of biostatistics, 2005

[94]: (PDF) Molecular Docking of the Inhibitory Activities of Selected Phytochemicals in Artemisia Afra Against NADH-Ubiquinone Oxidoreductase of Plasmodium Falciparum (PfNDH2)', ResearchGate, doi:10.11648/j.mc.20231101.12.

[95]: Chap I Syrup Formulation.

Annex 01: protox3 data bases

For the molecule tymol



Predicted LD50: 338mg/kg

Predicted Toxicity Class: 4

1 2 3 4 5 6

Average similarity: 100%

Prediction accuracy: 100%

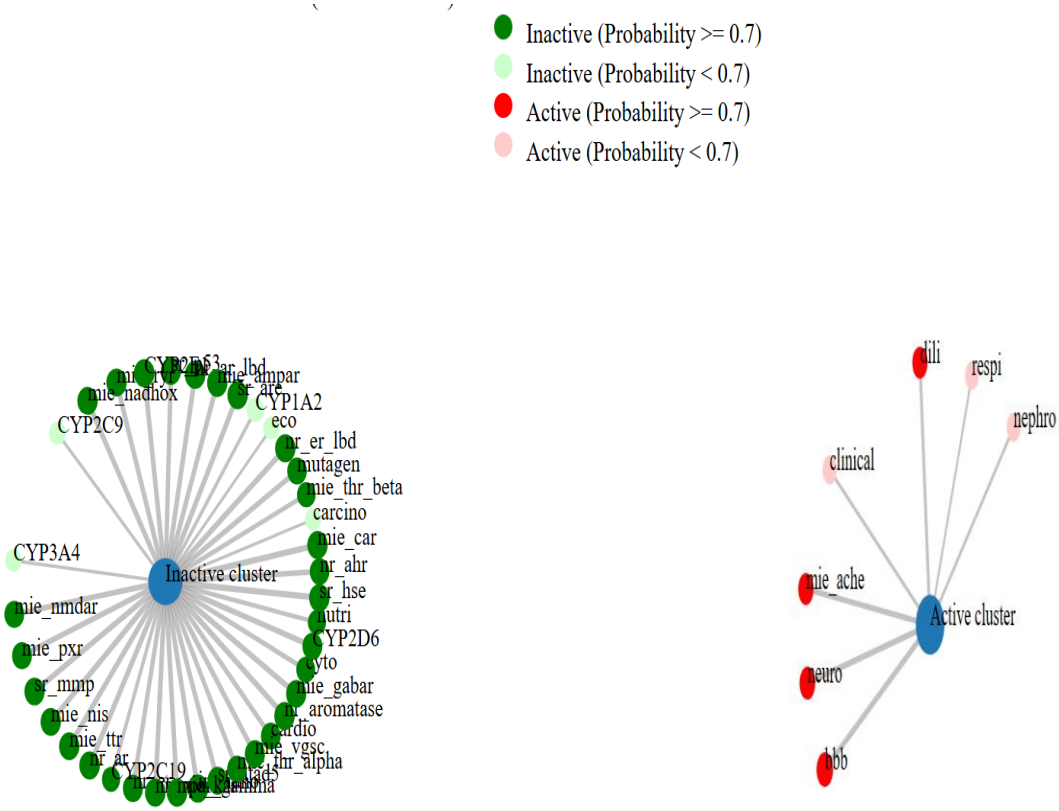
20% 40% 60% 80%

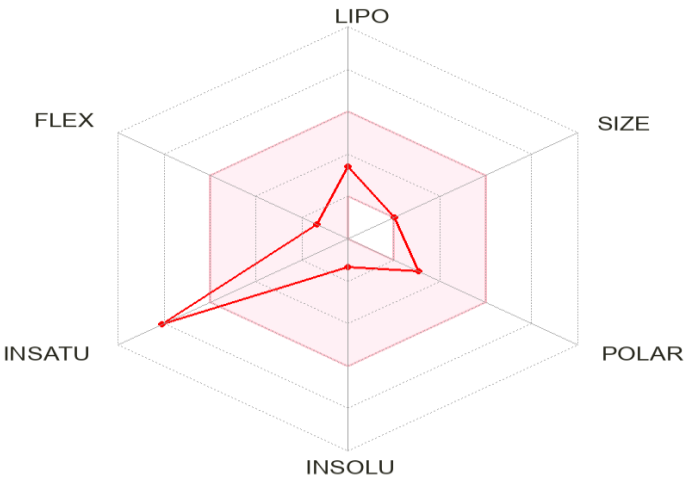
Print Toxicity Report

Name	CC(=O)NC1=CC=C(C=C
Molweight	151.16
Number of hydrogen bond acceptors	3
Number of hydrogen bond donors	2
Number of atoms	11
Number of bonds	11
Number of rotable bonds	2
Molecular refractivity	42.78
Topological Polar Surface Area	49.33
octanol/water partition coefficient(logP)	1.42

The toxicity radar chart is intended to quickly illustrate the confidence of positive toxicity results compared to the average of its class

The network chart


Inactive (Probability >= 0.7)
Inactive (Probability < 0.7)
Active (Probability >= 0.7)
Active (Probability < 0.7)
Inactive cluster
CYP3A4
CYP2C9
CYP1A2
nr_er_lbd
mutagen
mie_thr_beta
carcino
mie_car
nr_ahr
sr_hse
nutri
CYP2D6
cyto
mie_gabar
nr_aromatase
cardio
mie_vgsc
nr_thr_alpha
mie_nmdar
mie_pxr
sr_mmp
mie_nis
mie_ttr
nr_ar
mie_nadhox
sr_are
mie_ampar
eco
Active cluster
dili
respi
nephro
clinical
mie_ache
neuro
bbb


LIPO
SIZE
POLAR
INSOLU
INSATU
FLEX

Annex 02: Python Script for PIC50 Prediction from SMILES (SVR Model)

```
import numpy as np
import pandas as pd
from rdkit import Chem
from rdkit.Chem import Descriptors
from rdkit.ML.Descriptors import MoleculeDescriptors
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split, cross_val_predict
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_squared_error
from sklearn.feature_selection import VarianceThreshold, SelectKBest, f_regression
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import os
from sklearn.svm import SVR
#%%
# Configuration des graphiques pour publication scientifique
plt.style.use('seaborn-v0_8')
sns.set(style="whitegrid", font_scale=1.2)
plt.rcParams.update({
    'font.family': 'Times New Roman',
    'figure.dpi': 300,
    'savefig.dpi': 300,
    'axes.titlesize': 16,
    'axes.labelsize': 14,
    'xtick.labelsize': 12,
    'ytick.labelsize': 12
})
```